Econometrics and Statistics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Econometrics and Statistics

journal homepage: www.elsevier.com/locate/ecosta

Estimating a discrete distribution subject to random left-truncation with an application to structured finance

Jackson P. Lautier[a,*], Vladimir Pozdnyakov[b], Jun Yan[b]

[a] Department of Mathematical Sciences, Bentley University, Waltham, Massachusetts, USA
[b] Department of Statistics, University of Connecticut, Storrs, Connecticut, USA

ARTICLE INFO

Article history:
Received 26 February 2022
Revised 25 May 2023
Accepted 27 May 2023
Available online xxx

Keywords:
asset-backed security
asset-level disclosures
consumer lease securitization
product-limit estimator
reverse hazard rate
Reg AB II

ABSTRACT

Proper econometric analysis should be informed by data structure. Many forms of financial data are recorded in discrete-time and relate to products of a finite term. If the data is sampled from a financial trust, it will often be further subject to random left-truncation. The estimation of a distribution function from left-truncated data has been extensively addressed, but the case of discrete data over a known, finite number of possible values has not yet been thoroughly investigated. A precise discrete framework and suitable sampling procedure for the Woodroofe-type estimator for discrete data over a known, finite number of possible values is therefore established. Subsequently, the resulting vector of hazard rate estimators is proved to be asymptotically normal with independent components. Asymptotic normality of the survival function estimator is then established. Sister results for the left-truncating random variable are also proved. Taken together, the resulting joint vector of hazard rate estimates for the lifetime and left-truncation random variables is proved to be the maximum likelihood estimate of the parameters of the conditional joint lifetime and left-truncation distribution given the lifetime has not been left-truncated. A hypothesis test for the shape of the distribution function based on our asymptotic results is derived. Such a test is useful to formally assess the plausibility of the stationarity assumption in length-biased sampling. The finite sample performance of the estimators is investigated in a simulation study. Applicability of the theoretical results in an econometric setting is demonstrated with a subset of data from the Mercedes-Benz 2017-A securitized bond.

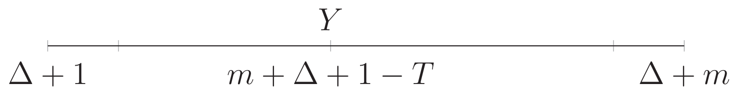
1. Introduction

The current outstanding issuance of consumer auto lease asset-backed securities (ABS) in the United States is nearly $35 billion (SIFMA, 2022), and the recent implementation of Reg AB II (U.S. Securities and Exchange Commission, 2016) has made a glut of public asset level ABS data available to investors for the first time. While more transparency into the underlying assets is generally a benefit for investors, the data may be difficult to analyze. This is because the legal structure of an ABS trust, the terms of a standard consumer automobile lease contract, and the nature of a monthly due date creates a need to consider left-truncation, a finite time horizon, and discrete-time, respectively, in estimating the distribution of consumer lease lifetimes.

* Corresponding author. Morison Hall, 175 Forest Street Waltham, MA 02452, USA
E-mail address: jlautier@bentley.edu (J.P. Lautier).

https://doi.org/10.1016/j.ecosta.2023.05.005
2452-3062/

(a) The variable T represents a random lease start time. Nonrandom time m is the origination time of the youngest lease in the trust as of the beginning of the trust observation window. Thus, $1 \leq T \leq m$, where $T, m \in \mathbb{N}$.



(b) Because we denote the time that the trust observation window begins by $\Delta + m$, nonrandom Δ denotes the minimum age of a lease in the trust as of time $\Delta + m$. Defining $Y = m + \Delta + 1 - T$ with $1 \leq T \leq m$ implies $\Delta + 1 \leq Y \leq \Delta + m$, where $Y, m, \Delta, T \in \mathbb{N}$.

(c) The random lease termination time, X, is observed if and only if $X \geq Y$. Nonrandom ω denotes the termination time of the lease with the longest active ongoing payments, and it coincides with the close of the trust observation window. Thus, $\Delta + 1 \leq X \leq \omega$, where $X, \Delta, \omega, Y \in \mathbb{N}$.

Fig. 1. The connected discrete random variables T, Y, X and the associated finite timelines for left-truncated data from an auto lease securitization.

To elaborate, consider an automotive lease securitization, such as Mercedes-Benz (2017), in which consumer automotive lease contracts are pooled together into a trust. Standard automotive lease contracts have a fixed and known duration, such as 36 months, with required monthly payments. Further, the payment performance of the lessee will be reported monthly, so the observed survival times of the lease contracts will be discrete within the nonnegative integers, $\mathbb{N}$. Left-truncation occurs because only those leases that remain active long enough to be collected into the trust will be observable by the investor. This is a form of bias under the general umbrella of *delayed entry* or *length-biased sampling* from the field of survival analysis (e.g., Asgharian et al., 2002; De Uña-Álvarez, 2004; Asgharian and Wolfson, 2005; Huang and Qin, 2011).

Formally, let X denote the random time of a lease contract termination (i.e., the lifetime or time-to-event random variable) and let T denote the random time of a lease contract origination. The context of our application restricts X and T to a known, finite subset of consecutive integers. If $\omega \in \mathbb{N}$ represents the age of the last lease termination in a sample, then $X \leq \omega$. Because issuers of structured debt typically have a legal obligation to the trust to select lease contracts with a minimum history of on-time payments, the youngest lease in the trust will have a minimum age of Δ as of the onset of the trust, where $\Delta \in \mathbb{N}$. Hence, each lease will have a minimum survival time of $\Delta + 1$, and so $\Delta + 1 \leq X \leq \omega$. If $m \in \mathbb{N}$ is the origination time of the youngest lease in the trust, then $1 \leq T \leq m$, and the trust starting time is $m + \Delta$. Our application implies the assumption $m + \Delta \leq \omega$. The integers Δ, m, and ω are non-random and known as of the onset of the problem. Observe that if we define $Y = m + \Delta + 1 - T$, then Y denotes a left-truncation random variable representing the minimum amount of time a lease must remain active to be observed in the trust. In other words, an investor will only observe those leases such that $X \geq Y$. For completeness, $\Delta + 1 \leq Y \leq \Delta + m$. We present a visualization of the connected random variables and timelines in Figure 1. Throughout, we assume X and T are independent, from which the independence of X and Y follows trivially.

The classical problem of estimating a distribution function in the presence of random left-truncation has been studied extensively. Specifically, if we consider two independent positive random variables X and Y with distribution functions F and G such that we only observe the pairs (X, Y) for which $Y \leq X$ and the pairs (X, Y) are assumed to be independent and identically distributed (i.i.d.), it is not difficult to find many thorough studies (e.g., Lynden-Bell, 1971; Woodroofe, 1985; Wang et al., 1986; Keiding and Gill, 1990; Stute, 1993; He and Yang, 1998a). The nature of securitization data requires us to further assume that X and Y are nonnegative integer-valued random variables with a known, finite number of possible values. The remaining assumptions of the classical problem remain valid.

This specialized discrete case of F and G has not yet been thoroughly studied, however. Two seminal related works are Woodroofe (1985) and Wang et al. (1986). Woodroofe (1985) proves consistency results for the Lynden-Bell (1971) estimator and shows its weak convergence to a Gaussian process but left the exact form of the covariance structure of the limiting process undefined. In deriving the asymptotic results, Woodroofe (1985) assumes continuous distribution functions F and G. Wang et al. (1986) extend the results of Woodroofe (1985) with a precise description of the asymptotic covariance structure.

It is noteworthy that this structure is the analogue of the covariance structure of the Kaplan–Meier estimator. Wang et al. (1986) allude to the idea that F and G need not be continuous in establishing strong consistency for the product limit estimator of F, but they assume continuity of F and G in working to define the covariance structure. Since Woodroofe (1985) and Wang et al. (1986), there have been many notable and significant contributions, many of which are summarized in Appendix A.

Typical approaches for avoiding an assumption of discrete-time may be problematic or inappropriate for consumer lease ABS data. First, a common approach is to force an assumption of continuous F and G, but this implies that ties are events with zero probability. Because there are likely many lease contracts with the same termination time, this creates an immediate complication. Second, it is possible to treat the lease performance data as interval-censored, where the event is assumed to occur within an interval of time (e.g., a month) but the exact time within the interval is unknown. With lease contracts and loan contracts more generally, however, payments made prior to a due date are treated the same as payments made on the due date (prepayments aside). In other words, a monthly payment is either received on-time or is delinquent. Thus, the ABS data is in actuality discrete with probability point masses at integer intervals; it is not a product of imprecise measurements. For similar reasons, even standard grouped survival data approaches (e.g., Prentice and Gloeckler, 1978) are not true representations of the failure time random variable for consumer lease data. One technical remark is that if the distribution function is known to contain discontinuities, but the location of such discontinuities is not known prior to performing the estimation, the analysis may be subject to additional complications, as in Rabhi and Asgharian (2017). We are working over a known, finite subset of $\mathbb{N}$, however, and thus may avoid this potentially cumbersome framework.

We now summarize our contributions. We thoroughly investigate a specialized discrete case of X and Y for the Woodroofe-type estimator. The first main result is that the vector of these estimators in discrete-time over a known, finite number of possible values is asymptotically normal with a fully-specified diagonal covariance matrix. The second main result is that the vector of Woodroofe-type estimates is the maximum likelihood estimate (MLE) for the parameters of the discrete bivariate distribution of (X, Y) given $Y \leq X$. We also find these results have a significant application potential for the large fixed-income asset class of consumer lease ABS data.

A detailed outline is as follows. In Section 2, we precisely define the joint conditional discrete sample space for X and Y, the related discrete conditional bivariate probability mass function and its connection to F and G through the hazard and *reverse hazard* rates, respectively, and a suitable sampling procedure to mimic the realities of securitized trust data. The sampling process we define differs from Woodroofe (1985) and is more suitable for ABS data. These preliminaries are necessary because the discrete case we examine has not before received a rigorous treatment. Section 2 closes by presenting the estimators and the first major result: all together, these estimates are the MLE of the discrete conditional bivariate distribution. Section 3 provides the next set of major results in that we state the asymptotic normality and independence of the estimation vector of the hazard rates for F and its analog for G; and, asymptotic normality of the estimator for the survival function of X and the estimator of the distribution function of Y. In all cases, the covariance matrix is completely specified. We also derive a hypothesis test for the shape of the distribution function with applications to length-biased sampling. In Section 4, we experimentally validate the results in Section 3 with a simulation study. In Section 5, we apply our results to a sample of data from the Mercedes Benz 2017-A securitized bond (Mercedes-Benz, 2017). We close with summary remarks in Section 6. Appendix A reviews related references, and Appendix B provides complete proofs of all major results.

2. Estimation

Section 2.1 briefly reviews notation and the identifiability results from Woodroofe (1985). We establish the specialized discrete sample space we consider in Section 2.2 (in the sequel, we may use the term *discrete* to refer to the more precise case of discrete F and G with known, finite support for convenience). This requires defining a discrete conditional bivariate probability mass function and related conditional distributions for discrete X and Y. Because of the discrete nature of X and Y, it is preferable to work in terms of the hazard rate of X and the reverse hazard rate of Y. For continuous F and G, it is common to instead work with the cumulative hazard function. We then connect the hazard and reverse hazard rates to F and G, respectively, in the discrete case. In Section 2.3, estimators for both the hazard and reverse hazard rates are then formally defined in the context of sampling from a left-truncated population rather than left-truncating a joint random sample. This is a further distinction from Woodroofe (1985). Finally, we formally state the result that the joint vector of estimates for the hazard and reverse hazard rates is a MLE for the parameters of the discrete conditional bivariate probability distribution for (X, Y) given $X \geq Y$.

2.1. Preliminaries

We use the notation of Woodroofe (1985). Let F and G be the distribution functions of non-negative independent random variables X and Y, respectively. Let H_* denote the joint distribution function of X and Y given $Y \leq X$, and let F_* and G_* denote the marginal distributions functions given $Y \leq X$ of X and Y, respectively. That is,

$$H_*(F, G, x, y) = \Pr(X \leq x, Y \leq y \mid X \geq Y),$$

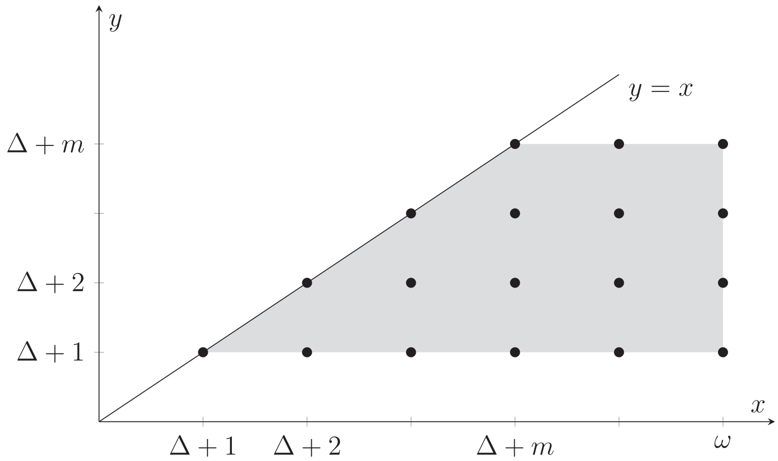


Fig. 2. The set of points on the plane with $(u, v) \in \mathbb{N}$ such that $u \in \{\Delta+1, \dots, \omega\}$, $v \in \{\Delta+1, \dots, \Delta+m\}$, and $v \le u$. The shaded region is the sample space of H_* and is denoted by trapezoid A. Because X and Y are discrete, all of the probability is contained in masses on the discrete points within the shaded region. If we assume that $\Pr(X = \Delta+1)$, $\Pr(Y = \Delta+1)$, $\Pr(X = \omega)$, and $\Pr(Y = \Delta+m)$ are strictly positive, then the edges of A are identifiable.

is the joint conditional distribution function with conditional marginal distributions F_* and G_*. We include F and G within the definition of H_* to emphasize which F and G are employed to construct H_*. For convenience, we may drop x and y from the notation for H_* when the meaning is clear or if the clarification is nonessential; i.e., $H_*(F, G)$.

We now review key observations made by Woodroofe (1985). Define

$$a_F = \inf\{z > 0 : F(z) > 0\} \ge 0,$$

and

$$b_F = \sup\{z > 0 : F(z) < 1\} \le \infty.$$

That is, (a_F, b_F) is the interior of the convex support of F and similarly (a_G, b_G) for G. To avoid complete left-truncation and full data loss, we must have $a_G < b_F$.

Next, we introduce two classes of distribution pairs (F, G). The first class includes all pairs of F and G that allow the construction of the two-dimensional distribution H_*,

$$\mathcal{K} = \{(F, G) : F(0) = 0 = G(0), \quad \Pr(Y \le X) > 0\}.$$

The second class includes only those pairs (F, G) that can be recovered from H_*,

$$\mathcal{K}_0 = \{(F, G) \in \mathcal{K} : a_G \le a_F, \quad b_G \le b_F\}.$$

Woodroofe (1985, Lemma 1) demonstrates that if we take any $(F, G) \in \mathcal{K}$ and let $F_0 = \Pr(X \le x \mid X \ge a_G)$ and $G_0 = \Pr(Y \le y \mid Y \le b_F)$, then $(F_0, G_0) \in \mathcal{K}_0$ and $H_*(F_0, G_0) = H_*(F, G)$. This subtle but important result implies that, if given H_*, we may not be able to recover the pair (F, G). This is because there is another pair, (F_0, G_0), that gives us exactly the same H_*. For example, in the context of our motivating problem, we only observe X when it is equal or greater than $\Delta+1$. Hence, it is impossible to get any information on the distribution of X over values less than $\Delta+1$.

2.2. Recovery

Woodroofe (1985, Theorem 1) shows that if we restrict our construction of H_* to the class $\mathcal{K}_0$, then this operation is *invertible*. More specifically, for every H_* based on some $(F, G) \in \mathcal{K}$ there is only one pair $(F_0, G_0) \in \mathcal{K}_0$ such that $H_*(F_0, G_0) = H_*(F, G)$ and this pair is given by F_0 and G_0. Moreover, Woodroofe (1985, Theorem 1) gives specific instructions on how to recover the cumulative hazard functions of F_0 and G_0: because there is a one-to-one relationship from the cumulative hazard function to the cdf, we immediately recover F_0 and G_0 as well.

We illustrate using the ABS framework of Section 1. Observe $F_0(x) = \Pr(\Delta+1 \le X \le x) / \Pr(X \ge \Delta+1)$ and $G_0(y) = \Pr(Y \le y) = G(y)$ because $\Delta+m \le \omega$ by assumption. The range of F_0 is $\{\Delta+1, \dots, \omega\}$, and the range of G_0 is $\{\Delta+1, \dots, \Delta+m\}$. Thus, from H_* based on the original F and G, it is possible to recover G but only the F_0 portion of F.

We further formalize the discrete case as follows. Let $X \in \mathbb{N}$ and $Y \in \mathbb{N}$ be independent random variables with ranges $\{\Delta+1, \dots, \omega\}$ and $\{\Delta+1, \dots, \Delta+m\}$, respectively. We will assume that $\Pr(X = \Delta+1)$, $\Pr(Y = \Delta+1)$, $\Pr(X = \omega)$, and $\Pr(Y = \Delta+m)$ are strictly positive, and $\Delta+m \le \omega$. Let A be a set of points on the plane $\mathbb{N} \times \mathbb{N}$ with integer-valued coordinates (u, v) such that $u \in \{\Delta+1, \dots, \omega\}$, $v \in \{\Delta+1, \dots, \Delta+m\}$, and $v \le u$. A visualization of A may be found in Figure 2.

Let

$$f(u) = \Pr(X = u), \quad g(v) = \Pr(Y = v), \quad \text{and} \quad \alpha = \Pr(Y \le X).$$

The bivariate distribution function H_* over the trapezoid A has probability mass function (pmf)

$$\begin{aligned} h_*(u,v) &= \Pr(X=u, Y=v \mid Y \le X) \\ &= \frac{f(u)g(v)}{\alpha}. \end{aligned} \tag{1}$$

The form of (1) implies there are distributions over A that cannot be a result of the random left-truncation procedure. The marginal distributions of H_* are given by

$$f_*(u) = \Pr(X=u \mid Y \le X) = \sum_v h_*(u,v),$$

and

$$g_*(v) = \Pr(Y=v \mid Y \le X) = \sum_u h_*(u,v).$$

To establish the recovery property, we must demonstrate it is possible to express the pmf f (or g) in terms of the pmf h_*. Woodroofe (1985, Theorem 1) proves it can be done by expressing the cumulative hazard rate function in terms of the joint cumulative distribution function (cdf) H_*. Because we deal only with discrete random variables, it is more convenient to demonstrate the same with the hazard rate for X,

$$\lambda(x) = \frac{\Pr(X=x)}{\Pr(X \ge x)},$$

where $x \in \{\Delta+1, \Delta+2, \ldots, \omega\}$. Specifically, we will show

$$\lambda(x) = \frac{f_*(x)}{C(x)}, \tag{2}$$

where

$$C(x) = \Pr(Y \le x \le X \mid Y \le X) = \sum_{v \le x \le u} h_*(u,v). \tag{3}$$

Observe,

$$\begin{aligned} C(x) &= \Pr(Y \le x \le X \mid Y \le X) \\ &= \frac{1}{\alpha}(\Pr(Y \le x) - \Pr(X < x, Y \le x)) \\ &= \frac{1}{\alpha}\Pr(Y \le x)\Pr(X \ge x). \end{aligned}$$

Hence,

$$\lambda(x) = \frac{\Pr(X=x)}{\Pr(X \ge x)} = \frac{\Pr(X=x, Y \le X)}{\Pr(Y \le X)} \frac{\Pr(Y \le X)}{\Pr(X \ge x)\Pr(Y \le x)} = \frac{f_*(x)}{C(x)}. \tag{4}$$

Because

$$C(x) \ge h_*(\omega, \Delta+1) = \frac{f(\omega)g(\Delta+1)}{\alpha} > 0,$$

for any x, we need not be concerned with $C(x)$ in the denominator of (4).

The bijection of the cdf F to the hazard rate λ is a standard result of survival analysis. For any integer x such that $\Delta+1 < x \le \omega$,

$$\begin{aligned} \prod_{\Delta+1 \le k < x} [1-\lambda(k)] &= \left[\frac{\Pr(X \ge \Delta+2)}{\Pr(X \ge \Delta+1)}\right]\left[\frac{\Pr(X \ge \Delta+3)}{\Pr(X \ge \Delta+2)}\right]\cdots\left[\frac{\Pr(X \ge x)}{\Pr(X \ge x-1)}\right] \\ &= \Pr(X \ge x), \end{aligned} \tag{5}$$

with the convention that (5) is unity for $x \le \Delta+1$. Because X is discrete, it is enough to know F at the points of discontinuity.

It is possible to derive an analog of formula (2) for the *reverse hazard rate* function (e.g., Block et al., 1998). The reverse hazard rate is the probability of the event of interest occurring in the current interval, given we know the event of interest occurred prior to the current interval. Formally, the reverse hazard rate, $\beta(y)$, is defined as

$$\beta(y) = \frac{\Pr(Y=y)}{\Pr(Y \le y)} = \frac{g_*(y)}{C(y)}, \tag{6}$$

where $y \in \{\Delta+1, \Delta+2, \ldots, \Delta+m\}$. Compare (6) with (2) to see the descriptive term *reverse* is natural. From (6), we have a one-to-one relationship for the cdf G that mirrors (5),

$$\Pr(Y \le y) = \prod_{\Delta+m \ge k > y} [1-\beta(k)], \tag{7}$$

where $\Delta+1 \le y \le \Delta+m$ and the convention that (7) is unity for $y \ge \Delta+m$.

2.3. Estimators

In the context of random left-truncation, different assumptions about the sampling procedure are possible. For example, Woodroofe (1985) assumes that there is a population of Xs and Ys, from which we take a sample of size N. Then we apply left-truncation to the sample, and this gives a sample of left-truncated pairs of *random* sample size n. Alternatively, we assume that there is the original population of Xs and Ys, and we apply left-truncation to the entire population. This results in a population of left-truncated pairs. Then we extract a sample of *deterministic* size n from the left-truncated population. In other words, our observations are directly from the distribution H_*. Given the practicalities of the securitization process, sampling from H_* directly is more appropriate for our application than the assumed sampling process of Woodroofe (1985). That is, our sampling process effectively samples from the already left-truncated lease data within the trust rather than imagines we are able to sit with the ABS issuer and see loans that did not meet the minimum survival requirements to be included in the trust. A theoretical divergence with generally limited practical significance, but its importance is evident with ABS data.

Formally, let $\{(X_i, Y_i)\}_{1\le i\le n}$ be i.i.d. pairs of random variables with distribution H_*. That is, $\{(X_i, Y_i)\}_{1\le i\le n}$ is a random sample of deterministic size n from H_* on trapezoid A of Figure 2. This is materially different than the sampling space of all possible target population pairs of (X, Y) absent left-truncation. In other words, referring again to Figure 2, there is a bias from the left-truncation condition in that some pairs, such as $u = \Delta + 1$ and $v = \Delta + m$, are not observable. This distinction warrants emphasis because it is erroneous to assume both pairs (X, Y) and (X_i, Y_i), $1 \le i \le n$, share the same properties. For example, while X and Y are assumed to be independent, X_i and Y_i clearly are not.

We desire to provide interval estimates for the hazard rates of F_0, the reverse hazard rates of G, and the cdfs F_0 and G from the i.i.d. sample $\{(X_i, Y_i)\}_{1\le i\le n}$ with distribution H_*. Examination of (2) indicates that the hazard rate $\lambda(x)$ is a ratio of two probabilities of events related to the random variables (X_i, Y_i), $1 \le i \le n$. Further, we have natural estimates of each probability within the ratio vis-à-vis the observed frequencies. This suggests the following intuitive estimator for the hazard rate,

$$\hat{\lambda}_n(x) = \frac{\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{X_i=x}}{\hat{C}_n(x)}, \tag{8}$$

where $\mathbf{1}_{(\cdot)}$ is the standard indicator function taking value 1 if statement $(\cdot)$ is true and 0 otherwise, and

$$\hat{C}_n(x) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}_{Y_j\le x\le X_j}. \tag{9}$$

From (5), an estimator for the cdf F_0 is immediate,

$$\hat{F}_n(x) = 1 - \prod_{\Delta+1\le k\le x}\left[1 - \hat{\lambda}_n(k)\right]. \tag{10}$$

Similarly, an estimator of the reverse hazard rate, $\beta(y)$, of Y follows,

$$\hat{\beta}_n(y) = \frac{\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{Y_i=y}}{\hat{C}_n(y)}, \tag{11}$$

as well as the cdf G,

$$\hat{G}_n(y) = \prod_{\Delta+m\ge k>y}\left[1 - \hat{\beta}_n(k)\right]. \tag{12}$$

It is theoretically satisfying that the estimators (10) and (12) coincide with the corresponding estimators (8) and (9) in Woodroofe (1985) despite the alternative construction required of our intended application to structured finance. This equivalence does not imply that the asymptotic results of Woodroofe (1985) directly follow for discrete F and G, however. This is because Woodroofe (1985) assumes a continuous F and G prior to proving its asymptotic results. This is a common assumption of many authors, which we summarize in Appendix A. The assumption of continuity is an extreme special case versus allowing F and G to take any form, and it helps avoid convergence argument complications caused by allowing for the possibility of discontinuities in the cdf. For an example of the complications in trying to allow for point masses without a priori knowledge of their locations, see the change point analysis and proofs of Rabhi and Asgharian (2017). Because the nature of ABS data will result in cdfs known to be discrete, we cannot utilize prior results of which continuity is used explicitly (e.g., Woodroofe, 1985, Theorem 2).

This is not to suggest that our forthcoming results allow for F and G to take any form. We consider another extreme special case, which is that F and G follow from a discrete distribution with a known, finite support. There still remains an attractive flexibility in that we impose limited restrictions on the shape of F and G within our framework. This flexibility is possible for two reasons. We know the locations of each probability point mass in F and G prior to the estimation, and we know there are only a finite number of them. Thus, we can treat the probability point masses as parameters to be estimated. It is this structure, which we formalized by h_* in (1), that allows us to handle the practical realities of ABS financial data with relative ease. For example, it is not any more difficult to handle ties in the X_is or Y_is among the observed sample $\{(X_i, Y_i)\}_{1\le i\le n}$ in proving the vector of estimates (8) and (11) are together the MLE of the parameters of the conditional bivariate distribution H_*. Compare this with Woodroofe (1985, pg. 168), for example, which assumes no ties for simplicity.

Remark. There are related cases of theoretical interest that we do not address. One such example is to relax the finite support restriction. That is, F and G are known to follow from a discrete random variable with known, countably infinite support (e.g., the positive integers). Such a case is not covered in this manuscript. For example, a parametric MLE requires a finite number of parameters (Mukhopadhyay, 2000, Definition 7.2.2), and so Theorem 2.1 does not cover the possibility of countably infinite support for F or G. We also present asymptotic results that converge to a fully specified multivariate normal distribution. A fully specified multivariate normal random vector requires a finite dimension (Ravishanker and Dey, 2002, Definition 5.2.2), and so we also do not address the case of discrete F and G with known, countably infinite support in any of the results in Section 3. Because addressing the possibility of countably infinite support for F and G requires very different tools to address than those of Appendix B and essentially all ABS are built from financial products with a finite term, this manuscript leaves the countably infinite case open and focuses on the unstudied case of discrete F and G with known, finite support. Nonetheless, we acknowledge the case of discrete F and G with known, countably infinite support is of theoretical interest, and we hope this Remark spur its future exploration.

It is noteworthy that the joint vector of estimates with components (8) and (11) can be shown to be the MLE of the parameters of the discrete conditional bivariate distribution H_*. The parameters of H_* are the discrete mass probabilities $0 \le f(u) \le 1$, $\Delta+1 \le u \le \omega$, and $0 \le g(v) \le 1$, $\Delta+1 \le u \le \Delta+m$ that comprise $h_*(u,v)$ in (1). Because $\sum_u f(u) = \sum_v g(v) = 1$, there are only $m+\omega-\Delta-2$ free parameters. There also exists an equivalent one-to-one parameterization of $h_*(u,v)$ using the hazard rates λ and β. Specifically, from (4) and (5),

$$f(u) = \lambda(u)\prod_{k=1}^{u-1}[1-\lambda(k)], \quad \text{and} \quad \lambda(u) = \frac{f(u)}{1-\sum_{k=1}^{u-1} f(k)}, \quad \Delta+1 \le u \le \omega, \tag{13}$$

with the conventions $\prod_{k=1}^{0}[1-\lambda(k)] = 1$ and $\sum_{k=1}^{0} f(k) = 0$. Similarly, from (6) and (7),

$$g(v) = \beta(v)\prod_{k=v+1}^{\Delta+m}[1-\beta(k)], \quad \text{and} \quad \beta(v) = \frac{g(v)}{1-\sum_{k=v+1}^{\Delta+m} g(k)}, \quad \Delta+1 \le v \le \Delta+m, \tag{14}$$

with the conventions $\prod_{k=\Delta+m+1}^{\Delta+m}[1-\beta(k)] = 1$ and $\sum_{k=\Delta+m+1}^{\Delta+m} g(k) = 0$. That there are still $m+\omega-\Delta-2$ free parameters is evident from the known hazard rate probabilities $\lambda(\omega) = \beta(\Delta+1) = 1$.

We now formally state the MLE property of the joint vector of estimates with components (8) and (11) in terms of the parameters of H_* in Theorem 2.1. The complete proof may be found in Appendix B.1. For ease of exposition, we also present an outline of the proof immediately following Theorem 2.1.

Theorem 2.1. *Define the discrete-mass trapezoid*

$$\mathcal{A} = \{(u,v) \in \mathbb{N} : \Delta+1 \le u \le \omega, \Delta+1 \le v \le \Delta+m, v \le u\},$$

and consider the bivariate distribution $h_*(u,v)$ defined in (1) over $\mathcal{A}$. Let $\{(X_i, Y_i)\}_{1\le i\le n}$ be n independent and identically distributed pairs of observations sampled from $h_*(u,v)$ such that $\hat{f}_{*,n}(u) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{X_i=u} > 0$ for $u \in \mathcal{A}$, and $\hat{g}_{*,n}(v) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{Y_i=v} > 0$ for $v \in \mathcal{A}$. Further define

$$\hat{\mathbf{\Lambda}}_n = \left(\hat{\lambda}_n(\Delta+1), \ldots, \hat{\lambda}_n(\omega-1), 1\right)^\top, \tag{15}$$

and

$$\hat{\mathbf{B}}_n = \left(1, \hat{\beta}_n(\Delta+2), \ldots, \hat{\beta}_n(\Delta+m)\right)^\top, \tag{16}$$

where $\hat{\lambda}_n$ and $\hat{\beta}_n$ follow from (8) and (11), respectively. Then the joint vector of estimates, $(\hat{\mathbf{\Lambda}}_n, \hat{\mathbf{B}}_n)^\top$, is a MLE for the parameters, $\lambda(u)$, $\beta(v)$, $u, v \in \mathcal{A}$, of the bivariate distribution $h_*(u,v)$.

Proof Outline. From the one-to-one correspondence of the two parameterizations of the distribution H_* and the invariance property of the MLE (e.g., Mukhopadhyay, 2000, Theorem 7.2.1, pg. 350), it is sufficient to find the MLE for the parameters f and g and demonstrate they are exactly equal to the estimates (8) and (11) in the same form as (13) and (14), respectively. It is preferable to define the likelihood in terms of the parameters f and g because the equivalent likelihood with a parameterization in terms of λ and β is cumbersome.

To maximize the likelihood, we restrict the parameter space of f and g to the convex set of all $0 < f, g < 1$ such that $\sum_u f(u) = \sum_v g(v) = 1$. The convexity of this restricted parameter space in conjunction with the behavior of the loglikelihood on the boundary of the parameter space confirms that the maximum point must lie within the restricted parameter space. We next solve the system of partial derivatives with respect to each parameter $f(u)$, $g(v)$ $u, v \in \mathcal{A}$, equated to zero sequentially to show the solution set admits only one solution, which must therefore be the global maximum and MLE. If we move sequentially from the minimum points of u and v, we can show the MLE for each f is exactly of the form (13). The complete result for g then follows by solving the system of partial derivative equations equated to zero sequentially from the maximum points of u and v (i.e., symmetry). □

There are some related results. For example, Vardi (1982) finds the MLE in the situation of a *length-biased distribution*, but the sampling procedure is not from h_*. Instead, two independent samples are used, the latter of which is from a length-biased distribution. In the derivation of Woodroofe (1985), the sampling procedure used therein is not from h_* but instead the complete pairs (X, Y), of which a random quantity are left-truncated whenever $Y > X$. Further, the results are given assuming no ties among the left-truncation or lifetime distribution observations. Wang (1987) also assumes the same sampling procedure as Woodroofe (1985) vis-à-vis stopping time theory. Keiding and Gill (1990) assume throughout that $\Pr(Y = X) = 0$, which is also not necessary in our framework. Furthermore, as indicated in Appendix A, the discrete case is largely left unstudied. We thus find our proof of Theorem 2.1 to be the first direct proof that the estimation vector with components (8) and (11) is indeed the MLE for the parameters of the discrete conditional distribution H_* over known, finite support.

3. Asymptotic Results

We now establish asymptotic normality of the hazard rate and reverse hazard rate estimators, along with the result of asymptotic independence. We also prove asymptotic normality of the survival function estimator for X and the distribution function estimator for Y. Finally, this section closes with a hypothesis test for the shape of G. All corresponding proofs may be found in Appendix B. Prior to these results, we provide three comments.

First, we review the assumptions of this section. The random variables X and Y are positive discrete random variables with known, finite support and distribution functions F and G, respectively. The random variables $\{(X_i, Y_i)\}_{1 \le i \le n}$ are independent and i.i.d. distributed pairs with distribution H_*. More specifically, $\{(X_i, Y_i)\}_{1 \le i \le n}$ are a random sample from a population with probability mass function h_* in (1), spanning the finite set of points on the plane with integer-valued coordinates (u, v) such that $u \in \{\Delta + 1, \dots, \omega\}$, $v \in \{\Delta + 1, \dots, \Delta + m\}$, $\Delta + m \le \omega$, and $v \le u$. Additionally, we will continue to assume that $\Pr(X = \Delta + 1)$, $\Pr(Y = \Delta + 1)$, $\Pr(X = \omega)$, and $\Pr(Y = \Delta + m)$ are strictly positive. Recall the support of h_* is summarized visually in Figure 2.

Second, we apply left-truncation to the entire population of X and Y, which yields a population of left-truncated pairs. From this left-truncated population, we draw a sample of deterministic size n. Therefore, in what follows, we investigate the limiting behavior as $n \to \infty$.

Third and finally, to state our asymptotic results it is convenient to introduce the following notation. Let (X_i, Y_i), $1 \le i \le n$ be a sample pair from h_*. Then we denote

$$\begin{aligned} c(u, v) &= \Pr\left(Y_i \le u \le X_i, Y_i \le v \le X_i\right) \\ &= \Pr\left(X \ge \max(u, v), Y \le \min(u, v) \mid Y \le X\right) \\ &= \sum_{y=\Delta+1}^{\min(u,v)} \sum_{x=\max(u,v)}^{L} h_*(x, y) \\ &= \frac{1}{\alpha} \Pr(Y \le \min(u, v)) \Pr(X \ge \max(u, v)). \end{aligned} \tag{17}$$

The various equation steps above (17) have been left as a form of summary of probability statements to show the connection between a single sampled pair (X_i, Y_i), $1 \le i \le n$, from h_*, the pair (X, Y) conditional on $Y \le X$, the pmf h_* itself, and the importance of assuming independence between X and Y. It is informative to observe $c(z, z) = C(z)$ and $c(u, v) = c(v, u)$.

3.1. Hazard and Reverse Hazard Rate

We first inspect the estimator $\hat{C}_n$ in the denominator of the hazard rate estimator in (8) with the multivariate Central Limit Theorem (CLT).

Lemma 1. ($\hat{\mathbf{C}}_n$ ***Asymptotic Normality)*** *Define* $\hat{\mathbf{C}}_n = (\hat{C}_n(\Delta + 1), \dots, \hat{C}_n(\omega))^\top$. *Then,*

$$\sqrt{n}(\hat{\mathbf{C}}_n - \mathbf{C}) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \boldsymbol{\Sigma}_c), \text{ as } n \to \infty,$$

where $\mathbf{C} = (C(\Delta + 1), \dots, C(\omega))^\top$ *and* $\boldsymbol{\Sigma}_c$ *is a covariance matrix* $\|\sigma_{k',k}\|$ *such that*

$$\sigma_{k',k} = \begin{cases} C(k)[1 - C(k)], & k' = k \\ c(k', k) - C(k')C(k), & k' \ne k \end{cases},$$

for $k', k = \Delta + 1, \Delta + 2, \dots, \omega$.

Lemma 2. *As* $n \to \infty$, $\hat{\mathbf{C}}_n \xrightarrow{\mathcal{P}} \mathbf{C}$.

The discrete nature of X and Y along with the known, finite sample space of trapezoid A yields attractive mathematical conveniences, which lead themselves naturally to computational programming. The same is true for the estimator of the hazard rate λ, which we now examine.

Theorem 3.1. *($\hat{\boldsymbol{\Lambda}}_n$ **Asymptotic Normality**) For $\hat{\boldsymbol{\Lambda}}_n$ defined in (15),*

$$\sqrt{n}(\hat{\boldsymbol{\Lambda}}_n - \boldsymbol{\Lambda}) \xrightarrow{\mathcal{L}} N\big(\mathbf{0}, \boldsymbol{\Sigma}_f\big), \text{ as } n \to \infty,$$

where $\boldsymbol{\Lambda} = (\lambda(\Delta+1), \lambda(\Delta+2), \ldots, \lambda(\omega))^\top$ *with* $\lambda(z) = f_*(z)/C(z)$ *and*

$$\boldsymbol{\Sigma}_f = \operatorname{diag}\left(\frac{f_*(\Delta+1)c(\Delta+1,\Delta+2)}{C(\Delta+1)^3}, \ldots, \frac{f_*(\omega-1)c(\omega-1,\omega)}{C(\omega-1)^3}, 0\right). \tag{18}$$

That is, the estimators $\hat{\lambda}_n(\Delta+1)$, …, $\hat{\lambda}_n(\omega)$ are asymptotically normal and independent.

Remark. There is an alternative form of $\boldsymbol{\Sigma}_f$ that may be preferable. Observe for $x \in \{\Delta+1, \ldots, \omega\}$,

$$\begin{aligned}\frac{f_*(x)c(x,x+1)}{C(x)^3} &= \frac{f_*(x)}{C(x)}\frac{\alpha^{-1}\Pr(Y \le x)\Pr(X \ge x+1)}{[\alpha^{-1}\Pr(Y \le x)\Pr(X \ge x)]^2}\\ &= \frac{\lambda(x)^2[1-\lambda(x)]}{f_*(x)}.\end{aligned}$$

Hence, alternatively,

$$\boldsymbol{\Sigma}_f = \operatorname{diag}\left(\frac{\lambda(\Delta+1)^2[1-\lambda(\Delta+1)]}{f_*(\Delta+1)}, \ldots, \frac{\lambda(\omega-1)^2[1-\lambda(\omega-1)]}{f_*(\omega-1)}, 0\right). \tag{19}$$

Further, (18) and (19) are equivalent when the true quantities for $f_*(\cdot)$, $c(\cdot,\cdot)$, and $C(\cdot)$ are replaced by their observed frequency estimators. That is, for $x \in \{\Delta+1, \ldots, \omega\}$ with

$$\hat{f}_{*,n}(x) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{X_i = x}, \quad \text{and} \quad \hat{c}_n(x,x+1) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{Y_i \le x, X_i \ge x+1}, \tag{20}$$

it is easy to show

$$\frac{\hat{f}_{*,n}(x)\hat{c}_n(x,x+1)}{\hat{C}_n(x)^3} = \frac{\hat{\lambda}_n(x)^2[1-\hat{\lambda}_n(x)]}{\hat{f}_{*,n}(x)}.$$

We now state the following corollary without proof for completeness.

Corollary 3.1.1. *As* $n \to \infty$, $\hat{\boldsymbol{\Lambda}}_n \xrightarrow{\mathcal{P}} \boldsymbol{\Lambda}$.

When estimating G is of interest, we may also obtain the sibling statement for the reverse hazard rate β as follows.

Theorem 3.2. *($\hat{\mathbf{B}}_n$ **Asymptotic Normality**) For $\hat{\mathbf{B}}_n$ defined in (16),*

$$\sqrt{n}(\hat{\mathbf{B}}_n - \mathbf{B}) \xrightarrow{\mathcal{L}} N\big(\mathbf{0}, \boldsymbol{\Sigma}_g\big), \text{ as } n \to \infty,$$

where $\mathbf{B} = (\beta(\Delta+1), \beta(\Delta+2), \ldots, \beta(\Delta+m))^\top$ *with* $\beta(z) = g_*(z)/C(z)$ *and*

$$\boldsymbol{\Sigma}_g = \operatorname{diag}\left(0, \frac{g_*(\Delta+2)c(\Delta+1,\Delta+2)}{C(\Delta+2)^3}, \ldots, \frac{g_*(\Delta+m)c(\Delta+m-1,\Delta+m)}{C(\Delta+m)^3}\right).$$

That is, the estimators $\hat{\beta}_n(\Delta+1), \ldots, \hat{\beta}_n(\Delta+m)$ are asymptotically normal and independent.

Remark. One may also write

$$\boldsymbol{\Sigma}_g = \operatorname{diag}\left(0, \frac{\beta(\Delta+2)^2[1-\beta(\Delta+2)]}{g_*(\Delta+2)}, \ldots, \frac{\beta(\Delta+m)^2[1-\beta(\Delta+m)]}{g_*(\Delta+m)}\right). \tag{21}$$

We state the following corollary without proof for completeness.

Corollary 3.2.1. *As* $n \to \infty$, $\hat{\mathbf{B}}_n \xrightarrow{\mathcal{P}} \mathbf{B}$.

3.2. Survival and Distribution Function

Often, a quantity of interest is the survival function, $S(x) = 1 - F(x)$. From (8) and (10), we have the estimator

$$\hat{S}_n(x) = \prod_{\Delta+1 \le k \le x} \left[1 - \hat{\lambda}_n(k)\right]. \tag{22}$$

Asymptotic normality also extends to (22), which we now show.

Theorem 3.3. ($\hat{\mathbf{S}}_n$ ***Asymptotic Normality)*** *For the estimator* $\hat{\mathbf{S}}_n = (\hat{S}_n(\Delta+1), \hat{S}_n(\Delta+2), \ldots, \hat{S}_n(\omega))^\top$,

$$\sqrt{n}(\hat{\mathbf{S}}_n - \mathbf{S}) \xrightarrow{\mathcal{L}} N\left(0, \mathbf{RK}\boldsymbol{\Sigma}_f[\mathbf{RK}]^\top\right), \text{ as } n \to \infty,$$

where $\mathbf{S} = (S(\Delta+1), S(\Delta+2), \ldots, S(\omega))^\top$, $\boldsymbol{\Sigma}_f$ *follows from Theorem 3.1,*

$$\mathbf{K} = \begin{bmatrix} -[1-\lambda(\Delta+1)]^{-1} & 0 & \ldots & 0 \\ -[1-\lambda(\Delta+1)]^{-1} & -[1-\lambda(\Delta+2)]^{-1} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -[1-\lambda(\Delta+1)]^{-1} & -[1-\lambda(\Delta+2)]^{-1} & \ldots & -[1-\lambda(\omega)]^{-1} \end{bmatrix},$$

and $\mathbf{R} = \mathrm{diag}(S(\Delta+1), S(\Delta+2), \ldots, S(\omega))$.

The sibling theorem to estimate G is as follows.

Theorem 3.4. ($\hat{\mathbf{G}}_n$ ***Asymptotic Normality)*** *For the estimator* $\hat{\mathbf{G}}_n = (\hat{G}_n(\Delta+1), \hat{G}_n(\Delta+2), \ldots, \hat{G}_n(\Delta+m))^\top$

$$\sqrt{n}(\hat{\mathbf{G}}_n - \mathbf{G}) \xrightarrow{\mathcal{L}} N\left(0, \mathbf{WM}\boldsymbol{\Sigma}_g[\mathbf{WM}]^\top\right), \text{ as } n \to \infty,$$

where $\mathbf{G} = (G(\Delta+1), G(\Delta+2), \ldots, G(\Delta+m))^\top$, $\boldsymbol{\Sigma}_g$ *follows from Theorem 3.2,*

$$\mathbf{M} = \begin{bmatrix} -[1-\beta(\Delta+1)]^{-1} & -[1-\beta(\Delta+2)]^{-1} & \ldots & -[1-\beta(\Delta+m)]^{-1} \\ 0 & -[1-\beta(\Delta+2)]^{-1} & \ldots & -[1-\beta(\Delta+m)]^{-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & -[1-\beta(\Delta+m)]^{-1} \end{bmatrix},$$

and $\mathbf{W} = \mathrm{diag}(G(\Delta+1), G(\Delta+2), \ldots, G(\Delta+m))$.

3.3. Hypothesis Test

In many applications, it is desirable to test if the distribution of F or G corresponds to a known distribution. An early reference is Hyde (1977). Additionally, Guilbaud (1988) generalizes the ordinary Kolmogorov–Smirnov one-sample tests based on the product-limit estimator. The test we develop from Theorem 3.5 is more akin to a goodness-of-fit test, however. In the following discussion of references, we use the broader terms “truncation” and “censoring”, as each reference may differ as to the specific type of incomplete data. Mandel and Betensky (2007) is related, though they assume continuous F and G to introduce several goodness-of-fit tests for the truncation distribution. Similarly, Hwang and Wang (2008) assume the lifetime, truncation, and censoring random variables are continuous in proposing a chi-square test for the hypothesis that the truncation distribution follows a parametric family. Further, the asymptotic properties of the nonparametric test of Ning et al. (2010) were derived assuming a continuous survival function. See also Moreira et al. (2014), in which goodness-of-fit tests are proposed for a semiparametric model under random double truncation. As there is no clear application to discrete F or discrete G with known, finite support, we propose a hypothesis testing procedure using a chi-square random variable. We state our results for the left-truncation distribution G.

Theorem 3.5. *Assume that G follows a known distribution over the discrete points* $\{\Delta+1, \ldots, \Delta+m\}$. *Then the test statistic*

$$\mathbb{Q}_G = \left[\sqrt{n}\left(\hat{\mathbf{B}}_n^* - \mathbf{B}^*\right)\right]^\top [\boldsymbol{\Sigma} g^*]^{-1} \left[\sqrt{n}\left(\hat{\mathbf{B}}_n^* - \mathbf{B}^*\right)\right] \xrightarrow{\mathcal{L}} \chi_q^2,$$

where $\hat{\mathbf{B}}_n^* = (\hat{\beta}_n(\Delta+2), \ldots, \hat{\beta}_n(\Delta+m))^\top$, $\mathbf{B}^* = \left(\beta(\Delta+2), \ldots, \beta(\Delta+m)\right)^\top$,

$$\boldsymbol{\Sigma}_g^* = \mathrm{diag}\left(\frac{\beta(\Delta+2)^2[1-\beta(\Delta+2)]}{g_*(\Delta+2)}, \ldots, \frac{\beta(\Delta+m)^2[1-\beta(\Delta+m)]}{g_*(\Delta+m)}\right),$$

and $q = \mathbf{card}\{\Delta+2, \ldots, \Delta+m\}$. *The point* $\Delta+1$ *with the degenerate estimator* $\hat{\beta}_n(\Delta+1) = 1$ *and* $Var[\hat{\beta}_n(\Delta+1)] = 0$ *is omitted from* $\mathbb{Q}_G$.

Specifically, it is often of interest to test if G follows a uniform distribution because it is an important assumption in the case length-biased sampling, see for instance Asgharian et al. (2002) and De Uña-Álvarez (2004). If the specialized discrete case of G within this manuscript is discrete uniform, it would allow us to simplify the parametric structure of h_*. There are previous investigations. For example, one stated use of the nonparametric maximum likelihood estimator (NPMLE) for the left-truncated distribution derived by Wang (1991) is to informally check the validity of the stationarity assumption. Similarly, Asgharian et al. (2006) propose a graphical method to check the stationarity of the underlying incidence times. Addona and Wolfson (2006) propose a formal test for stationarity of the incidence rate, but they require a continuous truncation density via Asgharian et al. (2006, Theorem 1). Our test, however, allows for exact p-value calculations and considers discrete G with known, finite support. Formally, Corollary 3.5.1 may be used to test if the left-truncation random variable follows a discrete uniform distribution.

Corollary 3.5.1. *Assuming the conditions and notation of Theorem 3.5, under the null hypothesis that G is a discrete uniform distribution over $\{\Delta+1,\dots,\Delta+m\}$, the test statistic*

$$\mathbb{Q}_U = \left[\sqrt{n}\left(\hat{\mathbf{B}}^*_n - \mathbf{B}^*_U\right)\right]^\top \left[\boldsymbol{\Sigma}^*_{g,U}\right]^{-1}\left[\sqrt{n}\left(\hat{\mathbf{B}}^*_n - \mathbf{B}^*_U\right)\right] \xrightarrow{\mathcal{L}} \chi^2_q,$$

where $\mathbf{B}^*_U = (1/2, 1/3, \dots, 1/m)$, and

$$\boldsymbol{\Sigma}^*_{g,U} = \operatorname{diag}\left(\frac{[1/2]^2[1-1/2]}{\hat{g}_{*,n}(\Delta+2)},\dots,\frac{[1/m]^2[1-1/m]}{\hat{g}_{*,n}(\Delta+m)}\right).$$

Consequently, for H_0 that Y is discrete uniformly distributed and significance level $0 \le \alpha \le 1$, one can reject H_0 if $\mathbb{Q}_U \le \chi^2_{q,\alpha/2}$ or $\mathbb{Q}_U \ge \chi^2_{q,1-\alpha/2}$, where $\chi^2_{q,\theta}$ is the $(100\times\theta)$th $(0<\theta<1)$ percentile of a chi-square distribution with q degrees of freedom. The accuracy of the asymptotic chi-square distribution was investigated for a discrete uniform G in a simulation study. The empirical distribution of the test statistics matches closely to the limiting chi-square distribution, which we validated down to a sample size of $n = 500$.

4. Simulation Study

In this section, we examine the finite sample behavior of the estimation vectors $\hat{\boldsymbol{\Lambda}}_n$ and $\hat{\mathbf{B}}_n$. In addition to serving as an experimental verification of Theorems 3.1 and 3.2, we investigate the minimum sample size of discrete-time left-truncated data needed to achieve a desired level of estimation accuracy. We proceed in two parts. First, for the purposes of illustration, we will consider a combination of classical distributions for the lifetime and left-truncation random variables. Second, the section will close with a combination of lifetime and left-truncation random variables that is a closer approximation to those observed within structured finance (e.g., Section 5).

Assume first that Y follows a discrete uniform distribution over $\mathcal{Y} = \{1, 2, \dots, 10\}$, and that X follows a truncated geometric distribution over $\mathcal{X} = \{1, 2, \dots, 24\}$. Specifically, the pmf of X is

$$\Pr(X = x) = \begin{cases} p(1-p)^{x-1}, & x = 1, 2, \dots, 23, \\ \sum_{x=24}^{\infty} p(1-p)^{x-1}, & x = 24, \\ 0, & \text{otherwise}, \end{cases} \tag{23}$$

where $0 < p < 1$. From this, we may calculate many key quantities of interest. For example, with $p = 0.20$,

$$\alpha = \Pr(Y \le X) = \sum_{y=1}^{10} \Pr(Y = y)\Pr(X \ge y) = 0.4463,$$

as well as the useful quantities (3), (17), (33), and (41). Observe that $\Delta = 0$, $m = 10$, and $\omega = 24$.

Consider the behavior of (18) across various values of p. For larger values of p, the variance of $\hat{\lambda}_n$ for values of X closer to 23 increases rapidly. This occurs because, as p increases, the probability of observing large values of X decreases. On the other hand, for very small values of p close to zero, the variance of $\hat{\lambda}_n$ for values of X close to 1 is the largest and rapidly decreases until $X = 10$, the final possible left-truncation point. This reconciles with the expected insights of (19): estimation accuracy of $\hat{\lambda}_n(x)$ is dependent on the quantities $\lambda(x)$ and $f_*(x)$.

In addition to demonstrating the asymptotic unbiasedness of the estimators (8) and (11), we will also compare the empirical covariance against the asymptotic covariance suggested by Theorems 3.1 and 3.2 by examining the resulting confidence interval estimates. It is desirable to keep the intervals within the unit interval. This can be done by constructing the confidence interval estimates on a log scale with the delta method (e.g., Mukhopadhyay, 2000, Theorem 5.3.5, pg. 261)

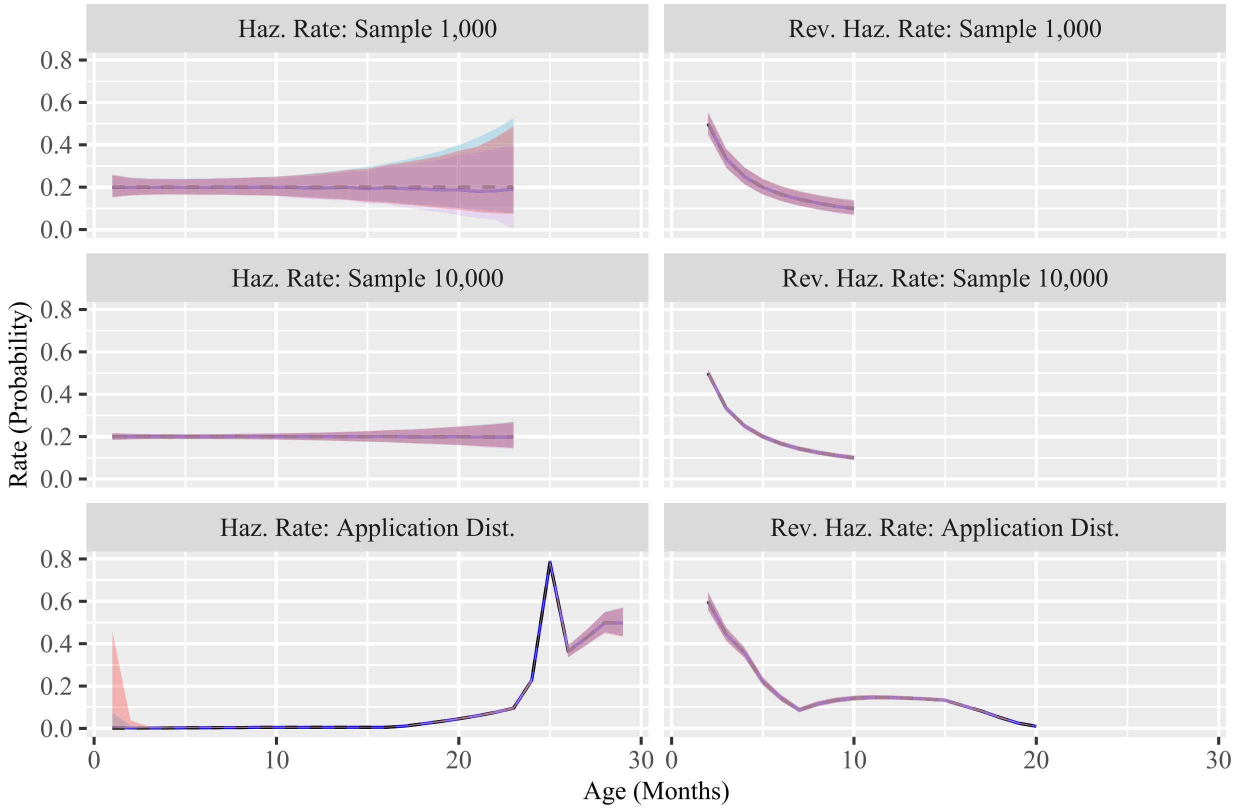


Fig. 3. A simulation verification of Theorems 3.1 (i.e., λ, left-column) and 3.2 (i.e., β, right-column) for sample sizes of $n = 1{,}000$ and $n = 10{,}000$ with classical distributions (the lifetime distribution follows a truncated geometric distribution at $x = 24$, as defined in (23) with $p = 0.20$ and the left-truncation distribution follows a discrete uniform distribution over the integers $\{1, \dots, 10\}$) and distributions more representative of an application to structured finance (compare Figures 4 and 5). Each chart presents a comparison of an average over all replicates of the estimate (blue lines) and true (dashed black lines), which are largely indistinguishable. Further the 95% true confidence intervals (blue ribbon), an average over all replicates of the 95% confidence intervals estimated from the empirical quantities (9) and (20) (red ribbon), and the middle 95th percentile (purple ribbon) of all the replicates each closely agree. The minor differences in the confidence intervals for the right tail of the lifetime distribution are eliminated as the sample size increases (row two versus row one). The only deterioration occurs in the very left tail of the lifetime distribution (bottom, left), which is a result of truncation causing very few simulated observations (the point estimate is still quite accurate). The results in the bottom row used a sample size of $n = 10{,}000$. All results used 1,000 replicates.

and then transforming them back exponentially to the original scale. For example, the 95% confidence intervals for $\lambda(x)$, $x \in \{1, \dots, 24\}$, are

$$\exp\left\{\ln \hat{\lambda}_n(x) \pm 1.96 \times \sqrt{\frac{1 - \hat{\lambda}_n(x)}{\hat{f}_{*,n}(x) \times n}}\right\}. \tag{24}$$

In our analysis summarized in the first two rows of Figure 3, we demonstrate the estimator's asymptotic unbiasedness and normality (we assume $p = 0.20$ and consider 1,000 replicates). For the asymptotic unbiasedness, we plot the true hazard and reverse hazard rates (dashed black lines) against an average of the 1,000 estimated replicates (blue lines). The two are largely indistinguishable, especially as n increases. For the asymptotic normality and covariance structure specified within Theorems 3.1 and 3.2, we compare three quantities. The first quantity is the true 95% confidence interval, and it is represented by the blue ribbon. The second quantity is the average of the estimated confidence intervals derived from the simulated data, i.e., (24), over the 1,000 replicates. We represent this quantity by the red ribbon. The third quantity is the middle 95th empirical percentile of the 1,000 replicates, and it is represented by the purple ribbon. All closely agree, especially as n increases. We also found that the off-diagonal elements in the empirical covariance calculation across the 1,000 replicates of all estimators are each very close to zero, which is further experimental validation of independence.

We next summarize the approximation accuracy across various sample sizes, and we observe it is a function of the underlying distribution. Intuitively, this can also be gleaned from (19); the variance of the estimator $\hat{\lambda}_n$ is a function of the

Table 1
Coverage percentages (CP) of 95% confidence intervals under various sample sizes in the simulation study adjusted for the frequency of unrealized simulations (UR). Top: $\lambda_n(x)$ for $x \in \{1, \dots, 24\}$; Bottom: $\beta_n(y)$, for $y \in \{1, \dots, 10\}$.

	$n = 250$		$n = 500$		$n = 750$		$n = 1,000$		$n = 10,000$	
x	CP	UR	CP	UR	CP	UR	CP	UR	CP	UR
1	95.9	0	94.4	0	93.8	0	93.8	0	95.6	0
2	96.2	0	94.5	0	94.5	0	95.4	0	94.9	0
3	95.3	0	94.5	0	95.0	0	95.7	0	95.8	0
4	95.3	0	94.6	0	95.0	0	94.3	0	93.9	0
5	95.4	0	94.1	0	95.7	0	94.5	0	95.8	0
6	93.8	0	94.5	0	93.9	0	94.1	0	95.8	0
7	93.7	0	94.4	0	95.6	0	95.6	0	95.4	0
8	94.5	0	95.3	0	95.7	0	95.2	0	95.2	0
9	95.3	0	95.1	0	95.8	0	94.4	0	93.0	0
10	95.1	0	95.5	0	95.9	0	96.7	0	93.3	0
11	95.2	0	95.4	0	94.4	0	94.4	0	93.8	0
12	95.9	0	95.6	0	95.3	0	95.7	0	95.1	0
13	94.6	1	94.9	0	94.5	0	95.7	0	96.2	0
14	96.0	0	96.3	0	95.7	0	94.6	0	94.6	0
15	95.5	6	94.6	0	94.0	0	95.3	0	95.7	0
16	95.1	23	96.8	2	93.7	0	94.5	0	94.4	0
17	95.5	37	96.6	1	95.7	0	95.2	0	95.1	0
18	94.5	77	96.0	5	96.8	1	95.0	0	94.8	0
19	93.1	131	94.7	21	96.1	3	95.2	0	94.8	0
20	92.6	204	95.0	43	95.3	4	95.3	2	95.8	0
21	91.1	296	95.3	69	95.1	20	95.6	1	95.0	0
22	90.0	347	93.6	146	94.3	49	95.7	20	94.7	0
23	87.5	431	91.3	206	93.8	84	95.5	35	94.8	0
	$n = 100$		$n = 250$		$n = 500$		$n = 1,000$		$n = 10,000$	
y	CP	UR	CP	UR	CP	UR	CP	UR	CP	UR
2	93.6	0	95.2	0	94.3	0	95.5	0	96.0	0
3	95.6	0	94.5	0	94.3	0	95.3	0	95.5	0
4	96.0	0	96.4	0	95.8	0	94.8	0	95.7	0
5	94.7	0	95.7	0	95.4	0	95.4	0	94.2	0
6	95.9	1	94.9	0	95.1	0	95.6	0	95.5	0
7	97.0	2	95.3	0	94.6	0	95.2	0	94.8	0
8	95.3	11	95.3	0	95.6	0	96.3	0	95.1	0
9	96.1	20	96.0	0	94.8	0	94.9	0	96.2	0
10	93.6	47	95.6	2	95.4	0	94.6	0	95.4	0

distributions of X and Y. Hence, we see that a larger sample size is necessary to control the approximation accuracy towards the right tail of the distribution of X, values of which occur with much smaller probability. We see minor tail failures of the approximation begin to materialize when n is as large as 1,000. On the other hand, the approximation for $\hat{\beta}_n$ still works well for $n = 1{,}000$. See Figure 3 for details.

Table 1 summarizes the observed coverage probability over the 1,000 replicates for various sample sizes, n. That is, the percentage of the 1,000 replicates of confidence intervals that contained the true value of $\lambda(x)$, $x \in \{1, \dots, 24\}$ and $\beta(y)$, $y \in \{1, \dots, 10\}$. We also track the number of replicates that did not return a valid estimate (i.e., we did not observe any samples of X or Y at a particular value). Given these results, we recommend that a practitioner use judgment and available references to estimate the probability of less frequent observations. The smaller these probabilities, the larger the sample to ensure the approximation works well. Alternatively, a practitioner may instead identify the values of X or Y that are of most interest. For example, the confidence interval approximation for $\hat{\lambda}_n$ still works very well for $X \leq 10$ when $n = 1,000$. More details may be found in Table 1. In addition, if reliable estimates of f_* and λ are available, then determining the appropriate sample size is only related to the method selected to set a simultaneous confidence region.

For the second part of our simulation study, we consider a combination of a lifetime random variable and a left-truncation random variable that is more representative of an application to structured finance; specifically, leases with an original termination schedule of 24 months. The probabilities are summarized in Figure 4. We can see the lifetime distribution obtains a peak near month 24, and the left-truncation distribution is not discrete uniform (compare with Figure 5). The bottom row of Figure 3 demonstrates experimental verification of Theorems 3.1 and 3.2. We see the true hazard rates and estimates overlap, which demonstrates asymptotic unbiasedness. Further, the true and empirical confidence intervals all closely agree, which demonstrates asymptotic normality and independence. The only deterioration in the estimator's asymptotic performance occurs with the confidence intervals at the very left tail of the lifetime distribution, which is a direct result of the combined lifetime and left-truncation random variable probabilities causing heavy left-truncation. The sample size for each of the 1,000 replicates was $n = 10{,}000$.

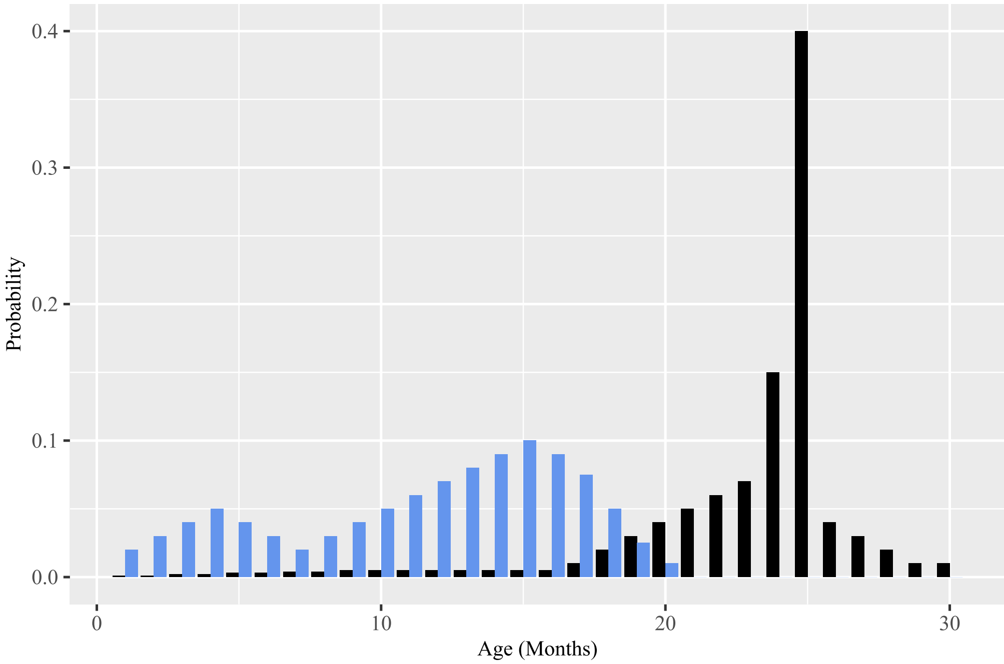


Fig. 4. The lifetime distribution (black bars) and left-truncation distribution (blue bars) more representative of an application to structured finance (compare with Figure 5) used to produce an additional simulation verification of Theorems 3.1 and 3.2 (bottom row, Figure 3).

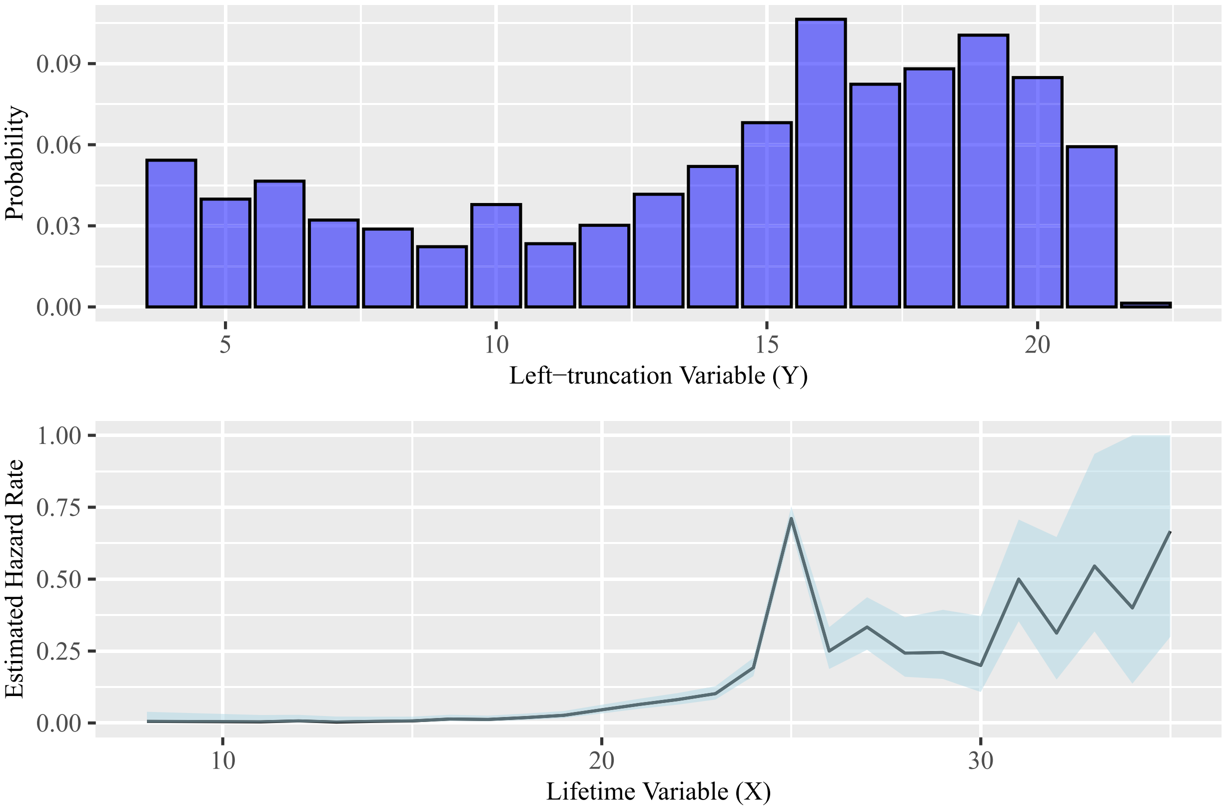


Fig. 5. Summary plots for $\hat{g}_{*,n}$ (top) and $\hat{\lambda}_n$ plus estimated 95% confidence intervals (bottom) for a subset of 24-month leases from the MBALT 2017-A securitization.

5. Application

To address the motivating problem in Section 1, we apply the estimation and asymptotic results of earlier sections to a subset of auto lease securitization trust data. Specifically, we examine the Mercedes–Benz Auto Lease Trust (MBALT) 2017-A financial transaction (Mercedes-Benz, 2017). Detailed data and performance records are available at the individualized contract level from the Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system, which is freely available to the public through the Securities and Exchange Commission (U.S. Securities and Exchange Commission, 2016). The MBALT 2017-A transaction has 56,402 lease contracts with original terms ranging from 24 to 60 months. For the purposes of illustration, we only consider ongoing lease contracts with an original termination schedule of 24 months. This reduces the sample to 866 lease contracts.

The MBALT 2017-A bond was placed in April of 2017. The transaction was paid in full and closed in August of 2019. Therefore, the observation window consists of 28 months. Monthly loan performance information is available on EDGAR. Lease contracts must be delinquent no more than 30 days to be included in the securitization trust (Mercedes-Benz, 2017). Hence, the lease contracts are all active as of the onset of the transaction. At initialization, the oldest lease in the trust was 21 months old, and the youngest lease was 3 months old. Thus, to use our notation, $\Delta = 3$ and $m = 18$. Though each lease is scheduled to terminate after 24 months, lease contracts may terminate early through default or consumer option. Additionally, lease contracts may extend beyond 24 months due to missed payments or various extension clauses. Therefore, to estimate the time of a lease termination, we searched the data for three consecutive months of a zero payment. Once three consecutive zero payments were found, the month of lease termination was assigned to be the month of the first zero. For example, if a lease contract recorded a zero payment for months 11, 12, and 13, then month 11 was assumed to be the lease termination age. After performing this search, we identified eight contracts that did not terminate during the observation window and were thus right-censored. For simplicity, we assume these eight leases all terminated as of the last observation month. (A related study, Lautier et al. (2023), generalizes the estimators of Section 2 to the case of right-censoring, see Section 6 for additional discussion.) The termination time of the oldest lease was 37 months, and so $\omega = 37$. Therefore, $Y \in \{4, \dots, 22\}$ and $X \in \{4, \dots, 37\}$. As a minor comment, we began counting T at 0 within this application, which is why the maximum bound of Y extends to $m + \Delta + 1 = 22$.

Table 2
Estimated distributions for the MBALT 2017-A application: the lifetime of interest (lease terminations, F_0) and the left-truncation random variable, G_0.

	F_0				G_0		
Age	$\hat{f}_{*,n}$	$\hat{\lambda}_n$	$s.e.[\hat{\lambda}_n]$	Age	$\hat{g}_{*,n}$	$\hat{\beta}_n$	$s.e.[\hat{\beta}_n]$
4	0	0	0	4	0.057	1	NA
5	0	0	0	5	0.042	0.424	1.577
6	0	0	0	6	0.048	0.331	1.229
7	0	0	0	7	0.033	0.186	0.917
8	0.001	0.005	0.161	8	0.030	0.143	0.763
9	0	0	0	9	0.023	0.100	0.621
10	0	0	0	10	0.039	0.145	0.675
11	0.001	0.004	0.115	11	0.024	0.082	0.505
12	0.002	0.007	0.147	12	0.031	0.096	0.516
13	0.001	0.003	0.093	13	0.043	0.117	0.531
14	0.002	0.006	0.115	14	0.053	0.127	0.515
15	0.003	0.007	0.121	15	0.069	0.143	0.502
16	0.008	0.014	0.152	16	0.107	0.182	0.503
17	0.008	0.012	0.135	17	0.082	0.124	0.404
18	0.014	0.019	0.157	18	0.087	0.117	0.373
19	0.022	0.027	0.177	19	0.097	0.118	0.355
20	0.040	0.046	0.223	20	0.080	0.090	0.305
21	0.058	0.065	0.260	21	0.053	0.059	0.250
22	0.068	0.081	0.298	22	0.001	0.001	0.041
23	0.079	0.102	0.345				
24	0.133	0.192	0.474				
25	0.397	0.711	0.607				
26	0.040	0.250	1.077				
27	0.040	0.333	1.354				
28	0.020	0.243	1.508				
29	0.015	0.245	1.739				
30	0.009	0.200	1.861				
31	0.018	0.500	2.601				
32	0.006	0.313	3.410				
33	0.007	0.545	4.418				
34	0.002	0.400	6.447				
35	0.002	0.667	8.009				
36	0	0	0				
37	0.001	1	NA				

In Figure 5, we plot the estimated hazard rate for 24-month leases within the MBALT 2017-A transaction. Most of these leases have terminated at month 25, which we would expect for a pool of leases contractually designed to terminate after 24 payments. There are a few additional observations. First, there is notable early lease termination activity beginning around lease age 20 months. Second, we have sporadic hazard rate behavior beyond lease age 25. Finally, the width of the 95% confidence band increases markedly beyond 25 months. The bands are quite narrow for leases that terminate prior to the original termination schedule of 24 months, however. Table 2 presents complete results for the estimated quantities $\hat{f}_{*,n}$, $\hat{\lambda}_n$, $\hat{g}_{*,n}$, and $\hat{\beta}_n$, along with the standard errors for $\hat{\lambda}_n$ and $\hat{\beta}_n$.

Additionally, it may be of interest to estimate the left-truncation random variable, Y. We present the estimated probability mass function for Y in Figure 5. An interested investor could use this information to recover T, the distribution of lease origination times. Information about T may be compared with economic trends or the "Selection of the Leases" section of Mercedes-Benz (2017), for example. Finally, it may be of interest to determine if the distribution of Y is discrete uniform to utilize a length-biased model (e.g., Asgharian et al., 2002; De Uña-Álvarez, 2004). Figure 5 is strong visual evidence that Y is not discrete uniform. We may test this formally using Corollary 3.5.1, which returns a statistic of $\mathbb{Q}_U = 1{,}530.6$. At $q = \mathbf{card}\{5, \ldots, 22\} = 18$ degrees of freedom, this corresponds to a p-value of effectively zero. Hence, we reject the null hypothesis of a discrete uniform distribution for Y. Rejecting the null in this case implies utilizing a method to estimate a distribution function for X that relies on the assumption that the left-truncation random variable is discrete uniform (i.e., stationarity), such as length-biased sampling, would be invalid for this application.

6. Discussion

The estimates in Table 2 and Figure 5 may have important applications to the automotive industry related to understanding the behavior of leaseholders when given the option to terminate or extend a lease contract. Financially, risk professionals can use our estimation procedures to model the relationship between consumer lessee behavior and the credit risk of securitized bonds. Automobile manufacturers may also have an interest in our application in terms of modeling the relationship between profitability and the structure of a consumer lease contract. The connective thread of this manuscript is that the estimates of Section 5 were produced using the theoretical results of Sections 2 and 3.

This is to our knowledge the first thorough exposition of the case of data subject to random left-truncation in the case of discrete X and Y with known, finite support. We proved that the random estimation vectors $\hat{\mathbf{\Lambda}}_n$ and $\hat{\mathbf{B}}_n$ are together an MLE for the parameters of the conditional bivariate distribution H_* and asymptotically normal with independent components (i.e., a diagonal covariance matrix). Both results utilized an alternative sampling and left-truncation framework from Woodroofe (1985), which was necessary to appropriately mimic the practicalities of consumer ABS data. We also further proved asymptotic normality extends to the survival function estimator $\hat{S}_n$ and the distribution function estimator $\hat{G}_n$. The last main result of this work was to establish a hypothesis test to examine the shape of the distribution of G, which has utility to formally test the stationarity assumption of the left-truncation distribution in length-biased sampling.

The practical realities of econometric data can inform statistical analysis, and we have identified a large group of securitized financial data that suggests the use of a survival analysis model adjusted for discrete-time data over a known, finite time horizon subject to random left-truncation. However, many forms of economic or financial data fall into the same criteria studied herein. Payment history is often recorded on a periodic basis, such as monthly, quarterly, or annually. For example, a monthly frequency is common for insurance products and debt instruments, such as insurance premiums, credit card payments, mortgages, auto loans, and so on. Further, many financial contracts typically have a fixed, finite term, such as any standard auto loan or term life insurance. Even whole life insurance, which is technically written with payments due in perpetuity may be instead interpreted as a fixed-length contract of unknown duration. In other words, it is reasonable to cap assumed policyholder lifetimes at an extreme age, such as 130 years. Our contributions related to the asymptotic statistical properties of the discrete distribution function estimators can be applied to and further investigated in alternative applications, such as those of insurance, mortgages, and other debt instruments.

We close with brief remarks on suggested generalizations and further work. Many applications of estimating a lifetime distribution random variable from observed data will also be subject to the further incomplete data complication of right-censoring. Because of the bias caused by left-truncation, it is important to determine if the censoring occurs before sampling with bias or afterwards. Tsai (2009, §3.3) provides a nice overview of the different potential censoring mechanisms of left-truncated data. For data sampled from ABS trusts, right-censoring occurs because the transaction eventually stops reporting loan performance data once the outstanding bond balance falls below the servicing costs. In other words, censoring occurs after sampling with bias. Lautier et al. (2023) is a thorough treatment that generalizes select theoretical results of this paper to consider right-censoring within the context of ABS. As such, Lautier et al. (2023) includes an extended financial pricing model and application to securitization data that utilizes the associated distribution estimators. In addition, general empirical economic analysis may benefit from the introduction of explanatory variables or covariates, similar to the classical regression models for survival data (e.g., Klein and Moeschberger, 2003, Section 2.6) but appropriately calibrated for the discrete-time setting of Section 2. We leave this problem open to further research. In addition, it is of theoretical interest to consider a discrete lifetime or left-truncation distribution over a known set of countably infinite values, such as the positive integers. Visually, this involves extending trapezoid A in Figure 2 upwards and to the right indefinitely. Because our application focuses on finite term financial products, however, we leave the discrete case with known countably infinite support open to further research.

Declaration of Competing Interest

The authors have no conflicts of interest to report.

Acknowledgments

Contents of this manuscript have benefited from seminar participants at the 2021 New England Statistical Symposium. We also thank the editorial team and two anonymous referees for providing substantive comments that greatly improved components of this manuscript. Jackson Lautier's work on this manuscript was supported in part by a National Science Foundation Graduate Research Fellowship under Grant No. DHE 1747453.

Appendix A. Literature Review

The following is a chronological review of results stemming from the seminal papers Woodroofe (1985) and Wang et al. (1986), which investigate the problem of estimating a distribution function from left-truncated data. Throughout this section, we denote the distribution functions of the random variable of interest, X, and the left-truncation random variable, Y, by F and G, respectively.

Chao and Lo (1988) further study the estimator of F by expressing a hazard process as i.i.d. means of random variables and imposing the same conditions as Woodroofe (1985). The result is the ability to represent the difference of F and its estimator as i.i.d. means of random variables to obtain weak convergence, including the associated covariance structures. Keiding and Gill (1990) reparametrize the left-truncation model as a three-state Markov process to invoke the statistical theory of counting processes by Aalen and Johansen (1978) to establish the NPMLE, consistency, asymptotic normality, and efficiency. Both papers derive results assuming continuity of F, however. Lai and Ying (1991) relax the continuity assumption of F in using martingale integral representations and empirical process theory to prove uniform strong consistency and weak convergence results, though they modify the product-limit estimator in doing so.

Gürler and Wang (1993) examine hazard functions and their derivatives for nonparametric kernel estimators. Similarly, they again assume continuity of G in proving asymptotic normality. Stute (1993) derives an almost sure representation of the estimator for F with weaker distributional assumptions than Woodroofe (1985) and improved error bounds. Chen et al. (1995) prove the Lynden-Bell (1971) estimator is uniformly strong consistent over the whole half line, a problem left open by Woodroofe (1985). Both papers assume continuity of F and G throughout. In part one of a two-part sequence, He and Yang (1998a) find a simpler representation for the estimator of the truncation probability to show strong consistency and asymptotic normality via an i.i.d. representation. While, these results are true for arbitrary F and G, they do not consider the estimators for the distribution functions for F and G. In part two, He and Yang (1998b) prove that the estimator for F obeys the strong law of large numbers when estimating F_0 for arbitrary and not necessarily continuous F (recall the distinction between F and F_0 in Section 2). This relaxes the assumption of continuity but does not address asymptotic normality. The classical problem of estimating F from truncated data also appears in textbooks (e.g., Karr, 1991; de la Peña and Giné, 1999; Owen, 2001; Hu, 2013), but any extended treatment assumes continuity of F (e.g., de la Peña and Giné, 1999, §5.5.3).

Finally, we expanded our review to consider the random left-truncation model along with right-censoring. A seminal work in this field is Tsai et al. (1987), which gives asymptotic results when left-truncated data are also subject to right-censoring. Nonetheless, the authors also assume continuous F. Similarly, the continuity of F and G is assumed in the related studies Uzonğullari and Wang (1992), Gijbels and Wang (1993), Gürler (1996), Zhou (1996), Zhou and Yip (1999), Asgharian and Wolfson (2005), and Huang and Qin (2011).

Appendix B. Complete Proofs

B1. Proof of Theorem 2.1

Proof. Without loss of generality, assume $\Delta = 0$. For convenience of notation, let $f_u \equiv f(u)$, $g_v \equiv g(v)$. Then, restating (1) in terms of the sampled pairs from h_*, (X_i, Y_i), $1 \le i \le n$, we have

$$h_*(u, v) = \Pr(X_i = u, Y_i = v) = \frac{f_u g_v}{\alpha}, \quad u, v \in \mathcal{A},$$

with the accompanying extended definition

$$\alpha = \Pr(Y \le X) = \sum_{u=1}^{\omega} f_u \left(\sum_{v=1}^{\min(u,m)} g_v \right) = \sum_{v=1}^{m} g_v \left(\sum_{u=v}^{\omega} f_u \right). \tag{25}$$

Therefore, the quantities $0 < f_u < 1$, $u \in \mathcal{A}$, and $0 < g_v < 1$, $v \in \mathcal{A}$ are the parameters to be estimated. Because we are working with a probability space, we must have $\sum_u f_u = \sum_v g_v = 1$. This implies there are $(\omega - 1) + (m - 1)$ free parameters.

Denoting $\boldsymbol{f} = (f_1, \dots, f_\omega)^\top$ and $\boldsymbol{g} = (g_1, \dots, g_m)^\top$, the likelihood and loglikelihood are then,

$$L(\boldsymbol{f}, \boldsymbol{g} \mid \{(X_i, Y_i)\}_{1 \le i \le n}) = \prod_{v=1}^{m} \prod_{u=v}^{\omega} \left[\frac{f(u)g(v)}{\alpha} \right]^{\sum_{i=1}^{n} \mathbf{1}_{(X_i, Y_i)=(u,v)}},$$

and

$$l(\boldsymbol{f}, \boldsymbol{g}) \equiv \frac{1}{n} \log L\big(\boldsymbol{f}, \boldsymbol{g} \mid \{(X_i, Y_i)\}_{1 \le i \le n}\big) = -\log \alpha + \sum_{v=1}^{m} \sum_{u=v}^{\omega} \hat{h}_{vu} \{\log f_u + \log g_v\}, \tag{26}$$

where

$$\hat{h}_{vu} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{(X_i, Y_i)=(u,v)}.$$

We desire to maximize (26). There are two ways to formulate this problem. The first is as a constrained optimization. Specifically, the parameter space of $\boldsymbol{f}$ and $\boldsymbol{g}$ is the $m \times \omega$ dimensional hypercube over the unit interval $\mathcal{I} = (0, 1)$, and we seek

$$\left\{ \max_{\boldsymbol{f}, \boldsymbol{g}} l(\boldsymbol{f}, \boldsymbol{g}) : \sum_{u=1}^{\omega} f_u = 1; \sum_{v=1}^{m} g_v = 1; \underset{1 \le u \le \omega}{f_u}, \underset{1 \le v \le m}{g_v} \in \mathcal{I} \right\}. \tag{27}$$

That is, $l(\boldsymbol{f}, \boldsymbol{g}) : (0, 1)^{m \times \omega} \mapsto \mathbb{R}$, subject to the constraints in (27). It is not straightforward to see that any solution will be a global maximum, however.

Alternatively, we can restrict the domain of $l(\boldsymbol{f}, \boldsymbol{g})$ to the convex set

$$\Psi = \left\{ \underset{1 \le u \le \omega}{f_u}, \underset{1 \le v \le m}{g_v} \in \mathcal{I} : \sum_{u=1}^{\omega} f_u = \sum_{v=1}^{m} g_v = 1 \right\}.$$

To see that Ψ is convex, without loss of generality, let $0 \le \varphi \le 1$ and suppose $f_u^* = \varphi f_u' + (1 - \varphi) f_u''$ for $f_u', f_u'' \in \Psi$ and $u \in \mathcal{A}$. Then

$$\begin{aligned} \sum_{u=1}^{\omega} f_u^* &= \sum_{u=1}^{\omega} \left\{ \varphi f_u' + (1 - \varphi) f_u'' \right\} \\ &= \varphi \sum_{u=1}^{\omega} f_u' + (1 - \varphi) \sum_{u=1}^{\omega} f_u'' \\ &= 1, \end{aligned}$$

and $f_u^* \in \Psi$. Thus, $l(\boldsymbol{f}, \boldsymbol{g}) : \Psi \mapsto \mathbb{R}$, and, from the convexity of Ψ, it is sufficient to claim we have found a global maximum if we can show $l(\boldsymbol{f}, \boldsymbol{g})$ has only one stationary point that is not on the boundary of Ψ.

A point on the boundary of Ψ implies that there exists at least one $f_u = 0$ or $g_v = 0$ for $u, v \in \mathcal{A}$. But, this immediately implies (26) explodes to negative infinity, (we assume here $\alpha > 0$ to avoid the degenerate case of complete data loss; see also the stricter conditions on $\hat{f}_{*,n}$ and $\hat{g}_{*,n}$ in the statement of Theorem 2.1). Hence, the maximum of (26) cannot lie on the boundary of Ψ, and, if we can show $l(\boldsymbol{f}, \boldsymbol{g})$ has only one stationary point, we can be assured it is a global maximum and therefore the MLE.

We now show the system of partial derivatives with respect to each parameter equated to zero has a single, unique solution. In the following, that $u, v \in \mathcal{A}$, i.e., $u, v \in \mathbb{N}$, is left assumed but will be dropped for ease of presentation. Observe first from (25),

$$\frac{\partial \alpha}{\partial f_u} = \sum_{v=1}^{\min(u,m)} g_v, \quad \text{and} \quad \frac{\partial \alpha}{\partial g_v} = \sum_{u=v}^{\omega} f_u.$$

Hence,

$$\frac{\partial l(\boldsymbol{f}, \boldsymbol{g})}{\partial g_v} = \frac{1}{g_v} \sum_{u=v}^{\omega} \hat{h}_{vu} - \frac{1}{\alpha} \frac{\partial \alpha}{\partial g_v} = 0, \quad 1 \le v \le m, \tag{28}$$

and

$$\frac{\partial l(\boldsymbol{f}, \boldsymbol{g})}{\partial f_u} = \frac{1}{f_u} \sum_{v=1}^{\min(u,m)} \hat{h}_{vu} - \frac{1}{\alpha} \frac{\partial \alpha}{\partial f_u} = 0, \quad 1 \le u \le \omega. \tag{29}$$

The simultaneous solution to (28) and (29) may be determined sequentially. We proceed by mathematical induction. That is, for $v = 1$, with (28),

$$\frac{1}{g_1} \sum_{u=1}^{\omega} \hat{h}_{1u} - \frac{1}{\alpha} \sum_{u=1}^{\omega} f_u = 0 \Rightarrow \hat{g}_1 = \alpha \sum_{u=1}^{\omega} \hat{h}_{1u} = \alpha \hat{C}_n(1).$$

Thus, for $u=1$, with (29),

$$\frac{1}{f_1}\sum_{v=1}^{1}\hat{h}_{v1}-\frac{1}{\alpha}\sum_{v=1}^{1}\hat{g}_v=0\Rightarrow \hat{f}_1=\frac{\hat{h}_{11}}{\hat{C}_n(1)}=\frac{\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{X_i=1}}{\hat{C}_n(1)}=\hat{\lambda}_n(1).$$

Consider now $v=2$ with (28),

$$\frac{1}{g_2}\sum_{u=2}^{\omega}\hat{h}_{2u}-\frac{1}{\alpha}\sum_{u=2}^{\omega}f_u=0\Rightarrow \hat{g}_2=\frac{\alpha\sum_{u=2}^{\omega}\hat{h}_{2u}}{1-\hat{f}_1}=\frac{\alpha\sum_{u=2}^{\omega}\hat{h}_{2u}}{1-\hat{\lambda}_n(1)}.$$

Thus, for $u=2$, with (29)

$$\begin{aligned}0&=\frac{1}{f_2}\sum_{v=1}^{2}\hat{h}_{v2}-\frac{1}{\alpha}\sum_{v=1}^{2}\hat{g}_v\\&=\frac{1}{f_2}\sum_{v=1}^{2}\hat{h}_{v2}-\frac{1}{\alpha}\left[\alpha\hat{C}_n(1)+\frac{\alpha\sum_{u=2}^{\omega}\hat{h}_{2u}}{1-\hat{\lambda}_n(1)}\right]\\&=\frac{1}{f_2}\sum_{v=1}^{2}\hat{h}_{v2}-\left[\frac{\hat{C}_n(1)-\sum_{i=1}^{n}\mathbf{1}_{X_i}+\sum_{u=2}^{\omega}\hat{h}_{2u}}{1-\hat{\lambda}_n(1)}\right]\\&=\frac{1}{f_2}\sum_{v=1}^{2}\hat{h}_{v2}-\frac{\hat{C}_n(2)}{1-\hat{\lambda}_n(1)}.\end{aligned}$$

That is,

$$\hat{f}_2=\hat{\lambda}_n(2)\left[1-\hat{\lambda}_n(1)\right].$$

Now assume the induction hypothesis for $1\le k<m$; i.e.,

$$\hat{g}_k=\frac{\alpha\sum_{u=k}^{\omega}\hat{h}_{ku}}{1-\sum_{j=1}^{k-1}\hat{f}_j},\qquad\text{and}\qquad \hat{f}_k=\hat{\lambda}_n(k)\prod_{1\le j<k}\left[1-\hat{\lambda}_n(j)\right],$$

with the conventions $\sum_{j=1}^{0}\hat{f}_j=0$ and $\prod_{1\le j<1}[1-\hat{\lambda}_n(j)]=1$. Then by (28),

$$\frac{1}{g_{k+1}}\sum_{u=k+1}^{\omega}\hat{h}_{k+1u}-\frac{1}{\alpha}\sum_{u=k+1}^{\omega}f_u\Rightarrow \hat{g}_{k+1}=\frac{\alpha\sum_{u=k+1}^{\omega}\hat{h}_{k+1u}}{1-\sum_{j=1}^{k}\hat{f}_j}.$$

But, for $1\le r\le k$,

$$\begin{aligned}1-\sum_{j=1}^{r}\hat{f}_j&=1-\hat{\lambda}_n(1)-\hat{\lambda}_n(2)[1-\hat{\lambda}_n(1)]-\cdots-\hat{\lambda}_n(r)[1-\hat{\lambda}_n(r-1)]\cdots[1-\hat{\lambda}_n(1)]\\&=\prod_{j=1}^{r}\left[1-\hat{\lambda}_n(j)\right].\end{aligned}\tag{30}$$

Thus,

$$\hat{g}_{k+1}=\frac{\alpha\sum_{u=k+1}^{\omega}\hat{h}_{k+1u}}{\prod_{j=1}^{k}[1-\hat{\lambda}_n(j)]}.$$

Therefore, for $u=k+1$, with (29),

$$\frac{1}{f_{k+1}}\sum_{v=1}^{k+1}\hat{h}_{vk+1}-\frac{1}{\alpha}\sum_{v=1}^{k+1}\hat{g}_v=0.$$

Further,

$$\begin{aligned}\sum_{v=1}^{k+1}\hat{g}_v&=\alpha\hat{C}_n(1)+\frac{\alpha\sum_{u=2}^{\omega}\hat{h}_{2u}}{1-\hat{\lambda}_n(1)}+\cdots+\frac{\alpha\sum_{u=k}^{\omega}\hat{h}_{ku}}{\prod_{j=1}^{k-1}[1-\hat{\lambda}_n(j)]}+\frac{\alpha\sum_{u=k+1}^{\omega}\hat{h}_{k+1u}}{\prod_{j=1}^{k}[1-\hat{\lambda}_n(j)]}\\&=\frac{\alpha\hat{C}_n(2)}{1-\hat{\lambda}_n(1)}+\frac{\alpha\sum_{u=3}^{\omega}\hat{h}_{3u}}{\prod_{j=1}^{2}[1-\hat{\lambda}_n(j)]}+\cdots+\frac{\alpha\sum_{u=k}^{\omega}\hat{h}_{ku}}{\prod_{j=1}^{k-1}[1-\hat{\lambda}_n(j)]}+\frac{\alpha\sum_{u=k+1}^{\omega}\hat{h}_{k+1u}}{\prod_{j=1}^{k}[1-\hat{\lambda}_n(j)]}\\&=\frac{\alpha\hat{C}_n(3)}{\prod_{j=1}^{2}[1-\hat{\lambda}_n(j)]}+\cdots+\frac{\alpha\sum_{u=k}^{\omega}\hat{h}_{ku}}{\prod_{j=1}^{k-1}[1-\hat{\lambda}_n(j)]}+\frac{\alpha\sum_{u=k+1}^{\omega}\hat{h}_{k+1u}}{\prod_{j=1}^{k}[1-\hat{\lambda}_n(j)]}\end{aligned}$$

$$\begin{aligned}
&\vdots\\
&= \frac{\alpha \hat{C}_n(k)}{\prod_{j=1}^{k-1}[1-\hat{\lambda}_n(j)]} + \frac{\alpha \sum_{u=k+1}^{\omega} \hat{h}_{k+1u}}{\prod_{j=1}^{k}[1-\hat{\lambda}_n(j)]}\\
&= \frac{\alpha \hat{C}_n(k+1)}{\prod_{j=1}^{k}[1-\hat{\lambda}_n(j)]}.
\end{aligned}$$

That is,

$$\hat{f}_{k+1} = \hat{\lambda}_n(k+1)\prod_{j=1}^{k}[1-\hat{\lambda}_n(j)].$$

Now, for $m < u \le \omega$,

$$\frac{1}{f_u}\sum_{v=1}^{m}\hat{h}_{vu} - \frac{1}{\alpha}\sum_{v=1}^{m}\hat{g}_v = 0 \iff \hat{f}_u = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{X_i=u}\frac{\prod_{j=1}^{m-1}[1-\hat{\lambda}_n(j)]}{\hat{C}_n(m)}.$$

Hence,

$$\begin{aligned}
\hat{f}_u &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{X_i=u}\frac{\prod_{j=1}^{m-1}[1-\hat{\lambda}_n(j)]}{\hat{C}_n(m)}\\
&= \frac{\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{X_i=u}}{\hat{C}_n(u)}\frac{\hat{C}_n(u)}{\hat{C}_n(u-1)}\frac{\hat{C}_n(u-1)}{\hat{C}_n(u-2)}\cdots\frac{\hat{C}_n(m+1)}{\hat{C}_n(m)}\prod_{j=1}^{m-1}[1-\hat{\lambda}_n(j)]\\
&= \hat{\lambda}_n(u)\left[\frac{\hat{C}_n(u-1)-\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{X_i=u-1}}{\hat{C}_n(u-1)}\right]\cdots\left[\frac{\hat{C}_n(m)-\sum_{i=1}^{n}\mathbf{1}_{X_i=m}}{\hat{C}_n(m)}\right]\prod_{j=1}^{m-1}[1-\hat{\lambda}_n(j)]\\
&= \hat{\lambda}_n(u)\prod_{j=1}^{u-1}[1-\hat{\lambda}_n(j)].
\end{aligned}$$

Lastly, because $\hat{\lambda}_n(\omega) = 1$,

$$\begin{aligned}
\sum_{u=1}^{\omega}\hat{f}_u &= \sum_{u=1}^{\omega}\left(\hat{\lambda}_n(u)\prod_{j=1}^{u-1}[1-\hat{\lambda}_n(j)]\right)\\
&= \hat{\lambda}_n(1) + (1-\hat{\lambda}_n(1))\sum_{u=2}^{\omega}\left(\hat{\lambda}_n(u)\prod_{j=2}^{u-1}[1-\hat{\lambda}_n(j)]\right)\\
&= \hat{\lambda}_n(1) + (1-\hat{\lambda}_n(1))[\cdots(1-\lambda_n(\omega-2))\left[\hat{\lambda}_n(\omega-1)+1-\hat{\lambda}_n(\omega-1)\right]\\
&\ \ \vdots\\
&= \hat{\lambda}_n(1) + 1 - \hat{\lambda}_n(1)\\
&= 1,
\end{aligned}$$

and the solution set $\hat{f}_u$, $u \in \mathcal{A}$, is in Ψ and omits only this single, unique solution. It is thus the global maximum of (26) and therefore the MLE. More specifically, we have found the MLE for the parameters f_u, $u \in \mathcal{A}$, and they are of the form (13). Therefore, $\hat{\mathbf{\Lambda}}_n$ is an MLE of f_u, for $u \in \mathcal{A}$ by the invariance property of the MLE (e.g., Mukhopadhyay, 2000, Theorem 7.2.1, pg. 350).

We can show $\hat{\mathbf{B}}_n$ is also a MLE for g_v, $v \in \mathcal{A}$, by moving sequentially from the other direction; e.g., for $m \le k \le \omega$, with (29),

$$\frac{1}{f_k}\sum_{v=1}^{m}\hat{h}_{vk} - \frac{1}{\alpha}\sum_{v=1}^{m}g_v = 0 \Rightarrow \hat{f}_k = \alpha\sum_{v=1}^{m}\hat{h}_{vk},$$

and thus, for $v = m$

$$\frac{1}{g_m}\sum_{u=m}^{\omega}\hat{h}_{um} - \frac{1}{\alpha}\sum_{u=m}^{\omega}\hat{f}_u = 0 \Rightarrow \hat{g}_m = \frac{\alpha\sum_{u=m}^{\omega}\hat{h}_{um}}{\sum_{u=m}^{\omega}\hat{f}_u} = \frac{\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{Y_i=m}}{\hat{C}_n(m)} = \hat{\beta}_n(m).$$

The remainder follows through symmetry. □

B2. Proof of Lemma 1

Proof. Observe

$$\hat{\mathbf{C}}_n = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{Y_i \le \Delta+1 \le X_i} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{Y_i \le \omega \le X_i} \end{bmatrix} = \frac{1}{n}\sum_{i=1}^{n} \begin{bmatrix} Y_{\Delta+1(i)} \\ \vdots \\ Y_{\omega(i)} \end{bmatrix}, \tag{31}$$

where $Y_{k(i)}$, $\Delta+1 \le k \le \omega$ are i.i.d. Bernoulli random variables with probability of success given by $\Pr(Y_i \le k \le X_i) = \Pr(Y \le k \le X \mid Y \le X) = C(k)$ for $k = \Delta+1, \ldots, \omega$. Thus, $\mathrm{E}[Y_{k(i)}] = C(k)$ and $\mathrm{Var}[Y_{k(i)}] = C(k)(1-C(k))$. Because

$$\mathbf{1}_{Y_i \le k' \le X_i}\mathbf{1}_{Y_i \le k \le X_i} = \mathbf{1}_{Y_i \le \min(k',k), X_i \ge \max(k'k)},$$

we have

$$\mathrm{E}[Y_{k'(i)}Y_{k(i)}] = \mathrm{E}[\mathbf{1}_{Y_i \le \min(k',k), X_i \ge \max(k'k)}] = c(k',k), \tag{32}$$

for $k', k = \Delta+1, \ldots, \omega$. Thus,

$$\begin{aligned} \mathrm{Cov}[Y_{k'(i)}, Y_{k(i)}] &= \mathrm{E}[Y_{k'(i)}Y_{k(i)}] - \mathrm{E}[Y_{k'(i)}]\mathrm{E}[Y_{k(i)}] \\ &= c(k',k) - C(k')C(k). \end{aligned}$$

Recall that (32) reduces to $C(k)$ when $k' = k$. The result then follows by the multivariate Central Limit Theorem (CLT) (Lehmann and Casella, 1998, Theorem 8.21, pg. 61). □

B3. Proof of Lemma 2

Proof. Applying the Weak Law of Large Numbers (Lehmann and Casella, 1998, Theorem 8.2, pg. 54-55) to (31) gives us the result. □

B4. Proof of Theorem 3.1

For convenience of notation, let

$$\begin{aligned} r(u,v) &= \Pr(X_i = \max(u,v), Y_i \le \min(u,v)) \\ &= \Pr(X = \max(u,v), Y \le \min(u,v) \mid Y \le X) \\ &= \sum_{y=\Delta+1}^{\min(u,v)} h(\max(u,v), y) \\ &= \frac{1}{\alpha}\Pr(X = \max(u,v))\Pr(Y \le \min(u,v)). \end{aligned} \tag{33}$$

Notice $r(z,z) = f_*(z)$ and $r(u,v) = r(v,u)$.

Proof. Recall (8)–(9) and observe

$$\hat{\mathbf{\Lambda}}_n - \mathbf{\Lambda} = \begin{bmatrix} \hat{\lambda}_n(\Delta+1) \\ \vdots \\ \hat{\lambda}_n(\omega) \end{bmatrix} - \begin{bmatrix} \lambda(\Delta+1) \\ \vdots \\ \lambda(\omega) \end{bmatrix} = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbf{1}_{X_i=\Delta+1}}{\hat{C}_n(\Delta+1)} - \frac{f_*(\Delta+1)}{C(\Delta+1)} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbf{1}_{X_i=\omega}}{\hat{C}_n(\omega)} - \frac{f_*(\omega)}{C(\omega)} \end{bmatrix}$$

$$= \mathbf{A}_n \times \frac{1}{n}\sum_{i=1}^{n} \begin{bmatrix} Z_{\Delta+1(i)} \\ \vdots \\ Z_{\omega(i)} \end{bmatrix},$$

where, for $\Delta+1 \le k \le \omega$,

$$Z_{k(i)} = \mathbf{1}_{X_i=k}C(k) - \mathbf{1}_{Y_i \le k \le X_i} f_*(k),$$

and $\mathbf{A}_n = \mathrm{diag}([\hat{C}_n(\Delta+1)C(\Delta+1)]^{-1}, \ldots, [\hat{C}_n(\omega)C(\omega)]^{-1})$. That is,

$$\hat{\mathbf{\Lambda}}_n - \mathbf{\Lambda} = \mathbf{A}_n \times \frac{1}{n}\sum_{i=1}^{n}\mathbf{Z}_{(i)},$$

where $\mathbf{Z}_{(i)} = (Z_{\Delta+1(i)}, \ldots, Z_{\omega(i)})^\top$, $1 \le i \le n$ are i.i.d. random vectors. We will also subsequently show that the components of random vector $\mathbf{Z}_{(i)}$ are uncorrelated.

More specifically, $\mathbf{1}_{X_i=x}$ is a Bernoulli random variable with probability of success $f_*(x)$ and, similarly, $\mathbf{1}_{Y_i\le x\le X_i}$ is a Bernoulli random variable with probability of success $C(x)$. Thus,

$$\mathrm{E}[Z_{k(i)}] = f_*(k)C(k) - C(k)f_*(k) = 0.$$

Therefore,

$$\mathrm{Cov}\Big[Z_{k(i)}, Z_{k'(i)}\Big] \tag{34}$$

$$\begin{aligned} &= E\Big[\Big(\mathbf{1}_{X_i=k}C(k) - \mathbf{1}_{Y_i\le k\le X_i}f_*(k)\Big)\Big(\mathbf{1}_{X_i=k'}C(k') - \mathbf{1}_{Y_i\le k'\le X_i}f_*(k')\Big)\Big] \\ &= C(k)C(k')\mathrm{E}\Big[\mathbf{1}_{X_i=k}\mathbf{1}_{X_i=k'}\Big] - f_*(k)C(k')\mathrm{E}\Big[\mathbf{1}_{X_i=k'}\mathbf{1}_{Y_i\le k\le X_i}\Big] \\ &\quad - C(k)f_*(k')\mathrm{E}\Big[\mathbf{1}_{X_i=k}\mathbf{1}_{Y_i\le k'\le X_i}\Big] + f_*(k)f_*(k')\mathrm{E}\Big[\mathbf{1}_{Y_i\le k\le X_i}\mathbf{1}_{Y_i\le k'\le X_i}\Big]. \end{aligned} \tag{35}$$

We proceed to calculate $\mathrm{Cov}[Z_{k(i)}, Z_{k'(i)}]$ by cases.

Case 1: $k = k'$.

Notice $\mathbf{1}_{X_i=k}\mathbf{1}_{X_i=k'} = \mathbf{1}_{X_i=k}$ and $\mathrm{E}[\mathbf{1}_{X_i=k}\mathbf{1}_{X_i=k'}] = f_*(k)$. Further,

$$\mathbf{1}_{X_i=k'}\mathbf{1}_{Y_i\le k\le X_i} = \mathbf{1}_{X_i=k, Y_i\le k\le X_i} = \mathbf{1}_{X_i=k}.$$

Hence, $\mathrm{E}[\mathbf{1}_{X_i=k}\mathbf{1}_{Y_i\le k'\le X_i}] = f_*(k)$. Also note that

$$\mathbf{1}_{Y_i\le k\le X_i}\mathbf{1}_{Y_i\le k'\le X_i} = \mathbf{1}_{Y_i\le k\le X_i},$$

and thus $\mathrm{E}[\mathbf{1}_{Y_i\le k\le X_i}] = C(k)$. Replacing the expectations in (35) yields

$$\begin{aligned} \mathrm{Cov}[Z_{k(i)}, Z_{k'(i)}] &= C(k)C(k')f_*(k) - f_*(k)C(k')f_*(k) \\ &\quad - C(k)f_*(k')f_*(k) + f_*(k)f_*(k')C(k) \\ &= C(k)^2 f_*(k) - 2f_*(k)^2C(k) + f_*(k)^2C(k) \\ &= f_*(k)C(k)[C(k) - f_*(k)]. \end{aligned} \tag{36}$$

However,

$$\begin{aligned} C(k) - f_*(k) &= \sum_{y=\Delta+1}^{k}\sum_{x=k}^{L} h_*(x,y) - \sum_{y=\Delta+1}^{k} h_*(k,y) \\ &= \sum_{y=\Delta+1}^{k}\sum_{x=k+1}^{L} h_*(x,y) \\ &= c(k, k+1). \end{aligned} \tag{37}$$

Replacing (37) in (36) we obtain $\mathrm{Cov}[Z_{k(i)}, Z_{k'(i)}] = f_*(k)C(k)c(k,k+1)$ when $k = k'$. We emphasize here that $c(\omega, \omega+1) = 0$.

Case 2: $k \ne k'$.

Certainly, $\mathbf{1}_{X_i=k}\mathbf{1}_{X_i=k'} = 0$ when $k \ne k'$. Therefore,

$$\mathrm{E}[\mathbf{1}_{X_i=k}\mathbf{1}_{X_i=k'}] = 0. \tag{38}$$

Assume $k < k'$ and notice $\mathbf{1}_{X_i=k'}\mathbf{1}_{Y_i\le k\le X_i} = \mathbf{1}_{X_i=k', Y_i\le k\le X_i}$. Thus, $\mathrm{E}[\mathbf{1}_{X_i=k'}\mathbf{1}_{Y_i\le k\le X_i}] = r(k', k)$. On the other hand, $\mathbf{1}_{X_i=k}\mathbf{1}_{Y_i\le k'\le X_i} = \mathbf{1}_{X_i=k, Y_i\le k'\le X_i} = 0$ because $\{X_i = k \cap k' \le X_i\} = \emptyset$ when $k < k'$. Now observe the symmetry between $\mathbf{1}_{X_i=k}\mathbf{1}_{Y_i\le k'\le X_i}$ and $\mathbf{1}_{X_i=k'}\mathbf{1}_{Y_i\le k\le X_i}$ to drop the assumption $k < k'$ and more generally claim

$$\begin{aligned} &- f_*(k)C(k')\mathrm{E}\Big[\mathbf{1}_{X_i=k'}\mathbf{1}_{Y_i\le k\le X_i}\Big] - C(k)f_*(k')\mathrm{E}\Big[\mathbf{1}_{X_i=k}\mathbf{1}_{Y_i\le k'\le X_i}\Big] \\ &\quad = -r(k,k')f_*(\min(k,k'))C(\max(k,k')). \end{aligned} \tag{39}$$

Further, $\mathbf{1}_{Y_i\le k\le X_i}\mathbf{1}_{Y_i\le k'\le X_i} = \mathbf{1}_{Y_i\le k\le X_i, Y_i\le k'\le X_i} = \mathbf{1}_{Y_i\le\min(k,k'), X_i\ge\max(k,k')}$. Hence,

$$\mathrm{E}[\mathbf{1}_{Y_i\le k\le X_i}\mathbf{1}_{Y_i\le k'\le X_i}] = c(k,k'). \tag{40}$$

Replacing the expectations (38), (39), and (40) in (35) and simplifying yields

$$\mathrm{E}[Z_{k(i)}Z_{k'(i)}] = f_*(\min(k,k')) \times \{f_*(\max(k,k'))c(k,k') - r(k,k')C(\max(k,k'))\}.$$

But,

$$\begin{aligned}
&= f_*(\max(k,k'))c(k,k') \\
&= \frac{\Pr(X=\max(k,k'),Y\le X)}{\alpha}\,\frac{\Pr(Y\le\min(k,k'))\Pr(X\ge\max(k,k'))}{\alpha} \\
&= \frac{\Pr(X=\max(k,k'))\Pr(Y\le\max(k,k'))}{\alpha}\,\frac{\Pr(Y\le\min(k,k'))\Pr(X\ge\max(k,k'))}{\alpha} \\
&= \frac{\Pr(X=\max(k,k'))\Pr(Y\le\min(k,k'))}{\alpha}\,\frac{\Pr(Y\le\max(k,k'))\Pr(X\ge\max(k,k'))}{\alpha} \\
&= r(k,k')C(\max(k,k')),
\end{aligned}$$

and so (35) is zero whenever $k\neq k'$. Therefore,

$$\mathrm{Cov}[Z_{k(i)},Z_{k'(i)}]=\mathrm{diag}\Big(f_*(\Delta+1)C(\Delta+1)c(\Delta+1,\Delta+2),\ldots,f_*(\omega)C(\omega)c(\omega,\omega+1)\Big)\equiv\mathbf{D}.$$

Now define

$$\bar{\mathbf{Z}}_n=\frac{1}{n}\sum_{i=1}^{n}\mathbf{Z}_{(i)},$$

and use the multivariate CLT (Lehmann and Casella, 1998, Theorem 8.21, pg. 61) to claim

$$\sqrt{n}[\bar{\mathbf{Z}}_n-\mathbf{0}]\xrightarrow{\mathcal{L}}N(\mathbf{0},\mathbf{D}),\ \text{ as } n\to\infty.$$

Further note by Lemma 2,

$$\mathbf{A}_n\xrightarrow{\mathcal{P}}\mathbf{V},\ \text{ as } n\to\infty$$

where $\mathbf{V}=\mathrm{diag}(C(\Delta+1)^{-2},\ldots,C(\omega)^{-2})$. Therefore, by multivariate Slutsky's Theorem (Lehmann, 1998, Theorem 5.1.6, pg. 283),

$$\sqrt{n}[\mathbf{A}_n\bar{\mathbf{Z}}_n]\xrightarrow{\mathcal{L}}N(\mathbf{0},\mathbf{V}\mathbf{D}\mathbf{V}^\top),\ \text{ as } n\to\infty.$$

Finally, observe $\mathbf{V}\mathbf{D}\mathbf{V}^\top=\mathbf{\Sigma}_f$ and $\mathbf{A}_n\bar{\mathbf{Z}}_n=\hat{\mathbf{\Lambda}}_n-\mathbf{\Lambda}$ to complete the proof. □

B5. Proof of Theorem 3.2

Proof. See the proof of Theorem 3.1, substituting g_* for f_* and adjusting the indicator logic as appropriate. It is useful to introduce similar notation to (33). That is,

$$\begin{aligned}
s(u,v)&=\Pr(Y_i=\min(u,v),X_i\ge\max(u,v)) \\
&=\frac{1}{\alpha}\Pr(Y=\min(u,v))\Pr(X\ge\max(u,v)).
\end{aligned}\tag{41}$$

□

B6. Proof of Theorem 3.3

Proof. Let $x\in\{\Delta+1,\ldots,\omega\}$ and recall (5) to write,

$$S(x)=\prod_{z=\Delta+1}^{x}[1-\lambda(z)].$$

Now consider the natural log,

$$\ln S(x)=\sum_{z=\Delta+1}^{x}\ln[1-\lambda(z)].$$

Hence,

$$\begin{aligned}
\sqrt{n}[\ln S_n(x)-\ln S(x)]&=\sqrt{n}\left[\sum_{z=\Delta+1}^{x}\ln\left(\frac{1-\lambda_n(z)}{1-\lambda(z)}\right)\right] \\
&=\sqrt{n}\left[\sum_{z=\Delta+1}^{x}\ln\left(1+\frac{\lambda(z)-\lambda_n(z)}{1-\lambda(z)}\right)\right].
\end{aligned}$$

But $\ln(1+x) = \sum_{n\geq 1}(-1)^{n+1}x^n/n$ and so

$$\sqrt{n}[\ln S_n(x) - \ln S(x)] = \sqrt{n}\left[\sum_{z=\Delta+1}^{x}\left\{\frac{\lambda(z)-\lambda_n(z)}{1-\lambda(z)} - \frac{1}{2}\left[\frac{(\lambda(z)-\lambda_n(z))^2}{(1-\lambda(z))^2}\right] + \cdots\right\}\right]$$
$$= \sqrt{n}\left[\sum_{z=\Delta+1}^{x}\frac{\lambda(z)-\lambda_n(z)}{1-\lambda(z)} + O_p(|\lambda(z)-\lambda_n(z)|^2)\right]$$
$$= \sqrt{n}\left(-\sum_{z=\Delta+1}^{x}\frac{\lambda_n(z)-\lambda(z)}{1-\lambda(z)}\right) + o_p(1), \quad (42)$$

where (42) follows by Corollary 3.1.1 and Slutsky's Theorem (Lehmann and Casella, 1998, Theorem 8.10, pg. 58). Now consider all $x \in \{\Delta+1, \ldots, \omega\}$ to write,

$$\sqrt{n}\begin{bmatrix}\{\ln S_n(\Delta+1) - \ln S(\Delta+1)\}\\ \vdots \\ \{\ln S_n(\omega) - \ln S(\omega)\}\end{bmatrix} = \mathbf{K} \times \sqrt{n}(\hat{\mathbf{\Lambda}}_n - \mathbf{\Lambda}) + o_p(1).$$

Thus, by Theorem 3.1 and multivariate Slutsky's Theorem (Lehmann, 1998, Theorem 5.1.6, pg. 283),

$$\mathbf{D} \times \sqrt{n}(\hat{\mathbf{\Lambda}}_n - \mathbf{\Lambda}) + o_p(1) \xrightarrow{\mathcal{L}} N(0, \mathbf{K}\mathbf{\Sigma}_f\mathbf{K}^\top), \text{ as } n \to \infty.$$

Finally, note $S(x) = \exp\{\ln S(x)\}$ and apply the multivariate delta method (Lehmann and Casella, 1998, Theorem 8.22, pg. 61) to complete the proof. □

B7. Proof of Theorem 3.4

Proof. Recall (12) and see the proof of Theorem 3.3. □

B8. Proof of Theorem 3.5

Proof. Begin with Theorem 3.2 along with (21) and use the following multivariate normal results: (1) all subsets of multivariate normal random vectors have themselves a normal distribution (Ravishanker and Dey, 2002, Result 5.2.8, pg. 154) and (2) a centered and scaled quadratic form of a p dimensional multivariate normal random vector is a chi-squared random variable with p degrees of freedom (Ravishanker and Dey, 2002, Result 5.3.3, pg. 167). The result then follows by the continuous mapping theorem (Lehmann and Casella, 1998, Corollary 8.11, pg. 58). □

B9. Proof of Corollary 3.5.1

Proof. By the Weak Law of Large Numbers (Lehmann and Casella, 1998, Theorem 8.2, pg. 54-55), $\hat{g}_{*,n} \xrightarrow{\mathcal{P}} g_*$. Further, if G is discrete uniform over $\{\Delta+1, \ldots, \Delta+m\}$, then for $y \in \{\Delta+1, \ldots, \Delta+m\}$

$$\beta(y) = \frac{\Pr(Y=y)}{\Pr(Y\leq y)} = \frac{1}{m}\frac{m}{y-(\Delta+1)+1} = \frac{1}{y-\Delta}.$$

Finally, use the results of Theorem 3.5 substituting $\beta(y)$ for $y \in \{\Delta+2, \ldots, \Delta+m\}$ as appropriate along with multivariate Slutsky's Theorem (Lehmann, 1998, Theorem 5.1.6, pg. 283) to complete the proof. □

References

Aalen, O.O., Johansen, S., 1978. An empirical transition matrix for non-homogeneous Markov chains based on censored observations. Scandinavian Journal of Statistics 5 (3), 141–150. http://www.jstor.org/stable/4615704

Addona, V., Wolfson, D.B., 2006. A formal test for the stationarity of the incidence rate using data from a prevalent cohort study with follow-up. Lifetime Data Analysis 12 (3), 267–284.

Asgharian, M., M'Lan, C.E., Wolfson, D.B., 2002. Length-biased sampling with right censoring. Journal of the American Statistical Association 97 (457), 201–209. doi:10.1198/016214502753479347.

Asgharian, M., Wolfson, D.B., 2005. Asymptotic behavior of the unconditional NPMLE of the length-biased survivor function from right censored prevalent cohort data. The Annals of Statistics 33 (5), 2109–2131. doi:10.1214/009053605000000372.

Asgharian, M., Wolfson, D.B., Zhang, X., 2006. Checking stationarity of the incidence rate using prevalent cohort survival data. Statistics in Medicine 25 (10), 1751–1767. doi:10.1002/sim.2326.

Block, H.W., Savits, T.H., Singh, H., 1998. The reversed hazard rate function. Probability in the Engineering and Informational Sciences 12 (1), 69–90. doi:10.1017/S0269964800005064.

Chao, M.-T., Lo, S.-H., 1988. Some representations of the nonparametric maximum likelihood estimators with truncated data. The Annals of Statistics 16 (2), 661–668. http://www.jstor.org/stable/2241747

Chen, K., Chao, M.-T., Lo, S.-H., 1995. On strong uniform consistency of the Lynden–Bell estimator for truncated data. The Annals of Statistics 23 (2), 440–449. http://www.jstor.org/stable/2242345
De Uña-Álvarez, J., 2004. Nonparametric estimation under length-biased sampling and Type I censoring: A moment based approach. Annals of the Institute of Statistical Mathematics 56 (4), 667–681. doi:10.1007/BF02506482.
Gijbels, I., Wang, J., 1993. Strong representations of the survival function estimator for truncated and censored data with applications. Journal of Multivariate Analysis 47 (2), 210–229. doi:10.1006/jmva.1993.1080.
Guilbaud, O., 1988. Exact Kolmogorov-type tests for left-truncated and/or right-censored data. Journal of the American Statistical Association 83 (401), 213–221. doi:10.1080/01621459.1988.10478589.
Gürler, Ü., 1996. Bivariate estimation with right-truncated data. Journal of the American Statistical Association 91 (435), 1152–1165. doi:10.1080/01621459.1996.10476985.
Gürler, Ü., Wang, J.-L., 1993. Nonparametric estimation of hazard functions and their derivatives under truncation model. Annals of the Institute of Statistical Mathematics 45 (2), 249–264.
He, S., Yang, G.L., 1998. Estimation of the truncation probability in the random truncation model. The Annals of Statistics 26 (3), 1011–1027. doi:10.1214/aos/1024691086.
He, S., Yang, G.L., 1998. The strong law under random truncation. The Annals of Statistics 26 (3), 992–1010. doi:10.1214/aos/1024691085.
Hu, C., 2013. Smoothing Spline ANOVA Models. Springer.
Huang, C.-Y., Qin, J., 2011. Nonparametric estimation for length-biased and right-censored data. Biometrika 98 (1), 177–186. http://www.jstor.org/stable/29777173
Hwang, Y.-T., Wang, C.-C., 2008. A goodness of fit test for left-truncated and right-censored data. Statistics & Probability Letters 78 (15), 2420–2425. doi:10.1016/j.spl.2008.02.035.
Hyde, J., 1977. Testing survival under right censoring and left truncation. Biometrika 64 (2), 225–230. doi:10.1093/biomet/64.2.225.
Karr, A.F., 1991. Point Processess and Their Statistical Inference. Marcel Dekker, Inc..
Keiding, N., Gill, R.D., 1990. Random truncation models and Markov processes. The Annals of Statistics 18 (2), 582–602. doi:10.1214/aos/1176347617.
Klein, J.P., Moeschberger, M.L., 2003. Survival Analysis: Techniques for Censored and Truncated Data, Second Edition. Springer.
Lai, T.L., Ying, Z., 1991. Estimating a distribution function with truncated and censored data. The Annals of Statistics 19 (1), 417–442. doi:10.1214/aos/1176347991.
Lautier, J.P., Pozdnyakov, V., Yan, J., 2023. Pricing time-to-event contingent cash flows: A discrete-time survival analysis approach. Insurance: Mathematics and Economics 110, 53–71. doi:10.1016/j.insmatheco.2023.02.003.
Lehmann, E., Casella, G., 1998. Theory of Point Estimation, 2nd Edition. Springer.
Lehmann, E.L., 1998. Elements of Large-Sample Theory. Springer.
Lynden-Bell, D., 1971. A method of allowing for known observational selection in small samples applied to 3CR quasars. Monthly Notices of the Royal Astronomical Society 155 (1), 95–118. doi:10.1093/mnras/155.1.95.
Mandel, M., Betensky, R.A., 2007. Testing goodness of fit of a uniform truncation model. Biometrics 63 (2), 405–412. doi:10.1111/j.1541-0420.2006.00710.x.
Mercedes-Benz, 2017. Prospectus: Mercedes-Benz Auto Lease Trust 2017-A. https://www.sec.gov/Archives/edgar/data/1537805/000114036117016403/form424b2.htm, Online; accessed 24 February 2022.
Moreira, C., De Uña-álvarez, J., Van Keilegom, I., 2014. Goodness-of-fit tests for a semiparametric model under random double truncation. Computational Statistics 29 (5), 1365–1379. https://doi.org/10.1007/s00180-014-0496-z
Mukhopadhyay, N., 2000. Probability and Statistical Inference. Marcel Dekker, New York, NY.
Ning, J., Qin, J., Shen, Y., 2010. Non-parametric tests for right-censored data with biased sampling. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72 (5), 609–630. doi:10.1111/j.1467-9868.2010.00742.x.
Owen, A.B., 2001. Empirical Likelihood. Chapman & Hall / CRC.
de la Peña, V.H., Giné, E., 1999. Decoupling: From Dependence to Independence. Springer.
Prentice, R.L., Gloeckler, L.A., 1978. Regression analysis of grouped survival data with application to breast cancer data. Biometrics 34 (1), 57–67. http://www.jstor.org/stable/2529588
Rabhi, Y., Asgharian, M., 2017. Inference under biased sampling and right censoring for a change point in the hazard function. Bernoulli 23 (4A), 2720–2745. doi:10.3150/16-BEJ825.
Ravishanker, N., Dey, D., 2002. A First Course in Linear Model Theory. Chapman & Hall (CRC).
SIFMA, 2022. US ABS securities: Issuance, trading volume, outstanding. https://www.sifma.org/resources/research/us-asset-backed-securities-statistics/. Online; accessed 24 February 2022.
Stute, W., 1993. Almost sure representations of the product-limit estimator for truncated data. The Annals of Statistics 21 (1), 146–156. doi:10.1214/aos/1176349019.
Tsai, W.Y., 2009. Pseudo-partial likelihood for proportional hazards models with biased-sampling data. Biometrika 96 (3), 601–615. doi:10.1093/biomet/asp026.
Tsai, W.-Y., Jewell, N.P., Wang, M.-C., 1987. A note on the product-limit estimator under right censoring and left truncation. Biometrika 74 (4), 883–886. doi:10.1093/biomet/74.4.883.
U.S. Securities and Exchange Commission, 2016. 17 CFR §229.1125 (Item 1125) Schedule AL — Asset-level information. https://www.govinfo.gov/app/details/CFR-2016-title17-vol3/CFR-2016-title17-vol3-sec229-1125 Online; accessed 24 February 2022.
Uzonğullari, Ü., Wang, J.-L., 1992. A comparison of hazard rate estimators for left truncated and right censored data. Biometrika 79 (2), 297–310. doi:10.1093/biomet/79.2.297.
Vardi, Y., 1982. Nonparametric estimation in the presence of length bias. The Annals of Statistics 10 (2), 616–620. http://www.jstor.org/stable/2240695
Wang, M.-C., 1987. Product limit estimates: A generalized maximum likelihood study. Communications in Statistics - Theory and Methods 16 (11), 3117–3132. doi:10.1080/03610928708829561.
Wang, M.-C., 1991. Nonparametric estimation from cross-sectional survival data. Journal of the American Statistical Association 86 (413), 130–143. http://www.jstor.org/stable/2289722
Wang, M.-C., Jewell, N.P., Tsai, W.-Y., 1986. Asymptotic properties of the product limit estimate under random truncation. The Annals of Statistics 14 (4), 1597–1605. http://www.jstor.org/stable/2241492
Woodroofe, M., 1985. Estimating a distribution function with truncated data. The Annals of Statistics 13 (1), 163–177. doi:10.1214/aos/1176346584.
Zhou, Y., 1996. A note on the TJW product-limit estimator for truncated and censored data. Statistics & Probability Letters 26 (4), 381–387. doi:10.1016/0167-7152(95)00035-6.
Zhou, Y., Yip, P.S.F., 1999. A strong representation of the product-limit estimator for left truncated and right censored data. Journal of Multivariate Analysis 69 (2), 261–280. doi:10.1006/jmva.1998.1806.